

This electronic thesis or dissertation has been downloaded from the King's Research Portal at <https://kclpure.kcl.ac.uk/portal/>



**The secrets of economic success and the causes of failure
the case of Imperial Rome**

Maher, George

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT



Unless another licence is stated on the immediately following page this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

The Secrets of Economic Success and the Causes of Failure

The Case of Imperial Rome

George P. M. Maher
King's College London
2015

Thesis submitted for the degree of Doctor of Philosophy

To my parents

ABSTRACT

The subject of this thesis is the economy of the Roman Empire from the beginning of the first through to the end of the third century AD. The approach taken is partly quantitative and draws on detailed quantitative research undertaken by scholars across a wide area. I concentrate the analysis on those few numbers which are likely to have the richest explanatory power and I consider the relationships between those numbers. The numbers analysed include the average amount of grain produced on a hectare of land, the size of the urban population, the average life expectancy and the silver content of the coinage. The approach also considers the institutions of imperial Rome, to the extent that they are relevant to economic achievement.

I consider the structure of the economy of the Roman Empire in terms of the separation of town and country, the agglomeration of individuals in the towns and the connections between those towns. I map out the development of the economy over time with a discussion of the creation and destruction of its currency.

I compare the performance of the economy of the Roman Empire with that of other economies by considering rural and urban productivity, trading activity, the infrastructure that supported trade and the knowledge base used in the production of goods and services.

The question that this thesis addresses is the following: How well did the economy of imperial Rome perform in terms of the quantity and quality of goods and services produced relative to the technology available and in comparison, with medieval and later economies and which institutions contributed to or hindered economic achievement?

I believe that the achievements of the Roman Empire in this period provide important lessons as to what is possible in economies generally and where some of the risks to economic success lie.

TABLE OF CONTENTS

1. Introduction	12
1.1. Purpose	12
1.2. Economic achievement defined	12
1.3. Previous approaches	12
1.4. My approach	16
1.4.1. Numbers	17
1.4.2. Relationships between numbers	19
1.5. The experiment	20
2. The agricultural sector and comparisons to other economies	21
2.1. Introduction	21
2.2. Seed productivity	22
2.2.1. Analysis of Roman seed productivity	22
2.2.2. Comparisons to other economies	31
2.3. Consumption	39
2.4. Labour structures	42
2.5. Conclusions	48
3. Levels of urbanisation and economic output	50
3.1. Introduction	50
3.2. The debate	50
3.3. Urban structures	51
3.4. The relationship between agricultural surplus and urbanisation	56
3.5. The relationship between the transport system and the size of towns	61
3.6. The relationship between urbanisation and total output	65
3.7. Productivity	69
3.8. Conclusions	78

4. The extent of trade and how it changed over time	81
4.1. Introduction	81
4.2. Previous approaches to trade in the economy of the Roman empire	81
4.3. Trade and Theory	85
4.3.1. Why goods are moved	85
4.3.2. Risks and costs which are a deterrent to trade	87
4.3.3. Intermediation	88
4.3.4. The prerequisites for trade	88
4.3.5. Scale of profits	89
4.4. What goods were moved and why	91
4.4.1. Wheat	91
4.4.2. Wine	92
4.4.3. Overall patterns	93
4.5. Risks and costs that restricted trade	93
4.5.1. Shipwreck and piracy	93
4.5.2. Inflation	95
4.5.3. Commercial risks	96
4.5.4. Costs	96
4.6. Intermediation	98
4.6.1. The Ship-Owner	98
4.6.2. Merchants	99
4.6.3. Trading Structures	101
4.7. Prerequisites	104
4.7.1. Information flow	104
4.7.2. State	104
4.7.3. Money	105
4.7.4. Port Structure	105
4.8. Conclusions	110

5. The expectation of life as a measure of economic progress	113
5.1. Introduction	113
5.2. The debate	113
5.3. The parameter P	116
5.4. Comparators and their implications	123
5.5. Relevant data and its interpretation	127
5.5.1. Inscription data	127
5.5.2. Ulpian's evidence	130
5.5.3. Data from Roman Britain	131
5.5.4. The Egyptian data	132
5.6. Applying P to population growth	134
5.7. Conclusions	136
6. Productive knowledge and the potential for growth	138
6.1. Introduction	138
6.2. The debate	138
6.3. Knowledge, labour and growth	140
6.4. Financial knowledge	143
6.5. Engineering knowledge	149
6.5.1. Tacit knowledge	149
6.5.2. Infrastructure development	151
6.6. Agricultural knowledge	153
6.7. Conclusions	156
7. The money system and its weaknesses	158
7.1. Introduction	158
7.2. Monetisation	158
7.3. The dynamics of the metallic currency	160
7.4. Relationship between the silver content of the denarius and inflation	167
7.5. Money supply – further features of this money system	168
7.5.1. State money creation	168
7.5.2. Private money creation	174
7.6. Inflation – the facts and the contributory factors	176
7.7. Conclusions	185

8. The system of state finances and fiscal incompetence	186
8.1. Introduction	186
8.2. The debate	187
8.3. Funding structures and flows	187
8.3.1. The continuing revenues and their importance	188
8.3.2. Other revenues	193
8.4. Capital expenditure	193
8.5. Current expenditure	196
8.5.1. Army manpower	197
8.5.2. Army pay accounts	197
8.5.3. Pay rates	200
8.6. State accounts	203
8.7. Conclusion	206
9. Conclusions	209
9.1. Introduction	209
9.2. Creation of the money system	209
9.3. Separation, agglomeration and connection	211
9.4. Destruction of the money system	214
9.5. The experiment	216
10. Appendices	218
11. Bibliography	228

LIST OF FIGURES

2.1 History of English grain yields	36
2.2 History of various European grain yields	37
3.1 Modern comparison of income and urbanisation levels	67
3.2 Total output as a multiple of rural output	68
4.1 Schematic illustration of the effect of profit retention on trade	90
5.1 Relationship between P and expectation of life: Model West	118
5.2 Relationship between P and expectation of life: All tables	119
5.3 Fit of British peerage data and English life tables	120
5.4 Relationship between P and expectation of life: British peerage	122
6.1 Modern wheat more productive than ancient wheat	155
7.1 Debasement of the coinage	171
7.2 State wheat prices	179
7.3 Private wheat prices	181
7.4 Private wine prices	182
7.5 Price of donkeys	183
7.6 Military pay	184
8.1 Schematic of current account revenue from Tribute	191
8.2 Illustration of variations in capital expenditure over time	194

LIST OF TABLES

1.1 Life expectancy and GDP compared	18
1.2 List of key parameters to explain economic performance	19
2.1 Summary of grain yields	29
2.2 The economic map	33
2.3 Example structure of agricultural unit	42
3.1 Distribution of town sizes	53
3.2 Illustration of structural changes in the English urban landscape	55
3.3 Relationship between the rural landscape and urbanisation	59
3.4 Relationship between land locked town size and hinterland radius	63
3.5 Comparison of land and sea supply	65
3.6 List of occupations	76
3.7 Relative earnings of different occupations	77
4.1 Shipwreck data	94
4.2 Freight charges for wheat	97
4.3 Change in processing capacity of principal ports around Rome	108
5.1 Illustration of the variety of mortality experiences	125
5.2 Illustration of a technique to estimate life expectancy	129
5.3 Example of the use of cemetery data	132
5.4 Difficulty of fitting Egyptian census data	133
5.5 Further difficulty of fitting Egyptian data	134
5.6 Relationship between births and intergenerational growth rates	135
7.1 Illustration of arbitrage opportunities	164
8.1 Summary of Capital Expenditure	195

8.2 Variations in army manpower over time	197
8.3 Example of soldiers pay accounts	198
8.4 Basic pay rates	201
8.5 Calculation of total cost of the army	204
8.6 Summary of scholarly estimates of total army costs	205
8.7 Schematic of State Current Account	206

ACKNOWLEDGEMENTS

I am grateful to Professor Dominic Rathbone for supervising this thesis, for his continuing encouragement, patience and helpfulness and especially for making this project so fascinating and enjoyable. I am grateful also to my examiners Professor Michael Crawford and Dr Claire Holleran for their comments and helpful discussions.

I would also like to acknowledge the help I received from Dr Christy Constantalopoulou, Dr Serafina Cuomo and Professor Catherine Edwards of Birkbeck who supervised my BA and MA essays and dissertation and especially the discussions I had with Dr Cuomo on knowledge and the economy.

Dr Andrew Hinde answered several emails and helped in my understanding of the history of British demography. The late Professor John Crook introduced me to Ulpian's valuation functions and kindly sent me some relevant passages from the Digest. Dr Kris Lockyear gave me useful background on Roman money and both Professor Henrik Mouritsen and Dr John Pearce commented on the early draft of my chapter on agriculture. In addition I had a number of conversations with Dr Pearce on grain productivity and on mortality. I had discussions with Dr Alexandra Sapoznik on medieval and early modern grain productivity and my fellow student Federico Ugolini and I discussed Roman port structures.

I had the benefit of insights into grain productivity from Patrick Mulcahy, a UK agronomist. Phil Skelton and Peter Seymour, both of ACE Insurance, gave me useful context on modern marine transportation. Adrian Gallup of the Government Actuary's Department provided helpful data extracted from the English Life Tables. Michael Townsend of the Institute of Historical Research helped me by identifying sources for population estimates of English towns. James McCormack, Sean McCormack and Martina Maher helped with the input of the over 1,500 data points on Slicher van Bath wheat yields. Helen Maher, Andy Staudt and Daniel Marx helped in the construction of some of the graphs and taught me how to make them on my own. Julian Leigh and Chris Waites read a later and an early version of the text respectively.

Finally I am very grateful for the support and continuing encouragement I received from Anne-Marie Harvey, Mitesh Rajput and Dr William Loschert.

CHAPTER 1: Introduction

1.1 Purpose

The purpose of this thesis is to present new ways of thinking about the economy of imperial Rome and to clarify the extent of the achievements of that economy. I believe that the economic structure that was developed in the first and second centuries AD, building on what had been previously constructed, was a major human achievement. Life expectancy was high, the level of agricultural productivity was higher than it was to be for many centuries afterwards and all of this was underpinned by trade for which the conditions were ideal. These achievements, however, were undermined by events in the third century AD.

1.2 Economic achievement defined

What is addressed in this thesis is economic output as measured by the quantity and types of goods and services produced, whether basic or highly developed. The greater the variety and the higher the quality, then the more developed the economy. The more secure the structures which underlie the flows that are essential to the production of these goods and services, then the stronger the economy. In this context I am interested in the institutions and technology of imperial Rome to the extent that they affected these outputs. I am also interested in comparisons with other economies to the extent that those comparisons help answer the question as to why the economy of imperial Rome produced more or better quality goods and services than another economy.

1.3 Previous approaches

A variety of approaches have been taken to the analysis of the imperial Roman economy. Previous approaches have included those which consider the institutions of the society and which draw economic conclusions from those considerations. Writing in 1926 Rostovtzeff analysed the Roman economy using such concepts as ‘capitalist’ and ‘bourgeois’ (or ‘middle class’) and explained the development of the economy in terms of, for example, the relationship between an aristocracy, an urban middle class and a lower class consisting of a rural peasantry and an urban proletariat. He believed that a strong middle class underlay the

strength of the economy of the Roman Empire.¹ Weber, however, believed that the ancient and modern economies differ fundamentally because of differences in the mentality in those economies, and that the use of class concepts was inappropriate. Weber believed that ancient economies were dominated by status groups which prevented the emergence of a market economy either because the members of these groups monopolised certain goods and services which were therefore not exchanged in the market place or these status groups prevented members from participating in particular activities. He believed that ancient status groups despised bargaining between the members of the status group and sometimes prevented the status group from bargaining with any party. This restricted their engagement in trade.² He located the emergence of modern economic behaviour for the first time in the Middle Ages. In particular, he saw the medieval town as being inhabited by a *homo oeconomicus* whereas in antiquity the polis had been inhabited by a *homo politicus*.³ Weber attributed the rise of modern capitalism to the emergence of a new work related mentality.⁴

Finley in writing about the ancient economy drew on the work of Weber and Polanyi and, like Weber, believed that activity in the ancient world was dominated by social rather than economic considerations. In *The Ancient Economy* he limits his consideration to the Graeco-Roman world which he characterised as being one of private ownership, private trade and private manufacture, and so he excludes from his analysis the near-eastern economies which were dominated from palace and temple complexes which owned most of the land, controlled trade and production centrally and organised life through a complex bureaucratic system.⁵

Finley took a pessimistic approach to the possibility of quantification.⁶ The work of Duncan-Jones stands in contrast and contains a very high level of numerical quantification. Duncan-Jones considered various aspects of the Roman economy such as agricultural profitability, price levels, urbanisation and demography but did not devote much attention to the examination of the relationships between the variables quantified and made no strong comparisons with other economies.⁷ Some of the very detailed analyses produced by Duncan-

¹ Rostovtzeff (1957).

² Weber (1978) 937.

³ Weber (1978) 1354.

⁴ Weber (1930).

⁵ Finley (1999) 28-9.

⁶ For example 'But first do we in fact know that productivity did not advance. Do we know anything about productivity at all? In a sense which can be expressed quantitatively the answer must be that we do not.' Finley (1965) 32.

⁷ Duncan-Jones (1982), Duncan-Jones (1990).

Jones, however, such as his reconstruction of soldiers' pay in the time of Diocletian support part of the quantitative structure I develop.

The approaches followed by Weber and Finley have been generalized to a broader consideration of how the Roman economy was shaped by its institutions, as well as by a consideration of transaction costs, such as those related to obtaining information necessary for trading decisions. These institutions are the sets of rules which provided the formal and informal constraints on economic interaction, which influenced motivation, which facilitated economic interaction through the reduction in uncertainty and which assisted economic actors in deciphering their environment.⁸ *The Cambridge Economic History of the Greco-Roman World*, for example, is one collaborative project which is influenced by these considerations and which contains an encyclopedic description of the economy of the Greco-Roman world including periods before and after that of this thesis. In particular it considers the economic effects of legal institutions, which I believe to be a significant contributor to the high level of productivity achieved in the time of the Roman empire.⁹

Other collaborative projects have been established, including the Oxford Roman Economy Project which was established in 2005 and is based in the Faculty of Classics at Oxford and has produced a range of working papers, publications and conferences. The Structural Determinants of Economic Performance in the Roman World, established in 2012, is a further network which organizes workshops and seminars and supports cooperation in research into economic performance in the Roman world. These sources have provided much material which is used in the following chapters.

Although these projects deepen knowledge of many detailed aspects of the Roman economy no framework has been produced which enables the debate to develop generally. My approach builds on the material produced in these projects and provides a framework in which the various detailed aspects can be understood together. My approach, while numerical in many aspects, fundamentally looks at human interaction, at what is achievable when individuals are agglomerated in centres of high population density and when those centres are highly connected. I work to discover which institutions, which sets of rules, matter by starting with an analysis of achievement and then moving from that to the identification of the institutional aspects that contribute to that achievement..

⁸ See North (1990) for a discussion of institutions and economic performance.

⁹ Frier and Kehoe (2007).

A successful framework depends on the appropriate selection of major categories for the analysis together with a deep analysis of the connections between the analytic categories. *The Cambridge Economic History of the Greco-Roman World* structures its analysis around the major categories of production, distribution and consumption, although with little cross referencing between these separate analyses.¹⁰ I believe that the analysis is likely to be more successful by alternatively being structured around the major categories of agriculture, urbanisation and trade and with connections being made between these separate analyses. High levels of urbanisation require high levels of agricultural surplus, for example, and high levels of productivity within urban centres require high levels of trade, as I demonstrate later. Connecting these analyses also allows data on agricultural productivity, for example, to inform views of urbanisation and the informational content of different data sets is more richly exploited.

Detailed quantitative analyses have been carried out by many scholars, and which cover many aspects of the Roman economy generally. For example, various literary texts which make reference to agricultural yields have been identified and compared. Accounts and records of estates in Roman Egypt have been reconstructed and additional data on agricultural yields has been extracted. Databases of medieval and later yields drawn from manorial and other records have been compiled. Archaeological work has provided information on the structure and size of agricultural units and archaeological data on the size of Roman towns has been compiled and estimates have been produced of the sizes of the populations which lived in the towns. Inscriptions have been analysed to form views on the occupations of urban dwellers. The construction and size of ships used in the time of imperial Rome have been researched. Information on port structures has been reconstructed from archaeological data and sizes estimated. The way in which traders and shippers organized themselves and the legal structures which supported their activities have been analysed. Sample data on the ages at death in Roman Egypt drawn from surviving census records has been compiled, as has age data drawn from funerary inscriptions and patterns of infant and adult commemoration have been researched. The details of particular engineering projects have been reconstructed from memorial inscriptions and archaeological evidence. The silver content of the coinage has been quantified in detail and the level of debasement mapped out over time. Price data has been reconstructed from papyrus documents found in Roman

¹⁰ There is, for example, extremely limited cross referencing by the authors of the chapters on production, consumption, distribution and the state to each other's earlier work and no cross referencing between these chapters.

Egyptian rubbish dumps. The amount of time spent in constructing aqueducts, temples, ports and other structures during the first century AD has been estimated, and estimates have also been made, based on building inscriptions, of the increase in the second century AD of this activity and its subsequent decline. The pay accounts of individual soldiers have been reconstructed as has the general level of pay. Further detailed analyses of different data types have been undertaken by scholars.¹¹

1.3 My approach

My aim is to use this and other data to understand the general structure of the economy of imperial Rome. The period covered is the first through the third centuries AD, and the territory covered is that of the Roman Empire, although I will draw on relevant data from earlier periods. This scope is more limited than that of Finley, which included the economy of the Roman Republic and also the economies of archaic, classical and later Greece, and even more limited than that of Weber which included medieval and later Europe. I will, however, from time to time draw comparisons between aspects of the economy of imperial Rome and other economies to understand the scale of the Roman achievement or failure and to test possible institutional reasons for those achievements or failures. Crucial to this process is the identification of aspects of the economy whose examination is likely to reveal the underlying structure and the selection of variables to review which are richly influenced by economic structures.

I believe that answering the question as to whether or not more or less by way of goods and services was available to a man who lived, for example, in London in the second century AD than to a man who lived there in the twelfth century or in the eighteenth century requires much more than a consideration of the institutions that then existed, although these institutions significantly determined what was produced. We can only understand how institutions work, or indeed if economically they matter at all, when we see what they actually do. The approach that I will take, therefore, will be formulated differently. I will examine the economic achievements and weaknesses of the imperial Roman economy and will then consider what institutions supported those achievements or undermined performance. Having answered these questions we can more clearly say whether or not the

¹¹ These investigations are described in detail in the following chapters. Examples of these contributions are Rathbone (1991), Rathbone (1996a), Rathbone (2000), Rathbone (2008), Rathbone and Temin (2008), Holleran (2012), Bang (2007), Wilson (2009), Wilson (2011), Temin (2004), Temin (2012), Scheidel (2011).

imperial Roman economy was different from medieval or early modern economies or a modern economy and in what respects.

1.4.1 Numbers

I will examine a range of parameters which enable the imperial Roman economy to be understood and placed alongside other economies. For example, I will consider how productive the agricultural sector was. The first parameter that I examine is grain productivity. I will concentrate on one grain type to simplify matters and make the comparisons easier, and I will express all output in the same terms of kilograms per hectare, again in order to facilitate comparison. This one parameter in itself is a powerful explanatory variable. Performance depends on local factors such as soil types and chance, such as year-on-year variations in weather. But it also depends on the application of human skill and care. As we will see, institutional factors, such as land holding structures, have a bearing on the grain productivity. Comparisons of the values of this parameter achieved under the imperial Roman economy with the values produced under other economic structures will enable conclusions to be drawn about the importance of the different institutions.¹²

I will consider how long people lived, that is life expectancy. This parameter directly reflects economic performance. For example, when access to food and warmth is more widespread and certain, then life expectancy is higher. Furthermore, life expectancy, measured as the average length of life of those who live within an economy, is affected by the distribution of wealth as much by the total wealth in an economy. In an economy where the benefits of increased output accrue only to a small part of the population, then the improvements in life expectancy will accrue only to that small part of the population and the effect on the average life expectancy will be limited. In an economy where benefits are more widely distributed, then the increase in average life expectancy for the same increase in total output will be greater. The following table summarises the different life expectancies that are experienced in the world today.¹³

¹² Andreau (2010) 46-8 criticises the use of models as part of historical research because they are not verifiable. I have used numbers very largely as a way of summarising facts. In cases where I have developed theories which summarise numerical relationships, such as in the chapter on life expectancy, I have tested theory against data.

¹³ <https://www.cia.gov/library/publications/resources/the-world-factbook/> Retrieved 29 October 2015. Life expectancy rounded to the nearest year and GDP per capita to the nearest US\$ 1,000.

Table 1.1 Life expectancy and GDP compared

1: Country	2: Life Expectancy	3: Per Capita GDP
United Kingdom	83 years	US\$ 40,000
United States	82 years	US\$ 55,000
Saudi Arabia	77 years	US\$ 52,000
Russia	77 years	US\$ 25,000
China	75 years	US\$ 13,000
India	69 years	US\$ 6,000
South Africa	64 years	US\$ 13,000
Zimbabwe	58 years	US\$ 2,000

Some knowledge acquired in the nineteenth century and after is now widespread, such as the knowledge of the harmful effects of dirty water and of what spreads infection, and, although they are not as readily available in some parts of the world, modern drugs can be found in all economies.

The economic and institutional structures that apply in these different countries vary considerably. Power structures vary between different forms of democracy, an autocratic monarchy and a kleptocracy. Levels of equality, in terms of the distribution of wealth, also vary. Property ownership rights vary across these countries. In Zimbabwe there is state interference in the private enjoyment of property. In Saudi Arabia the national resources are owned by the ruling family, but part of the income from those assets is widely distributed and there is advanced healthcare. Private ownership of land is highly developed and protected in the United States and the United Kingdom. The countries in this sample are further differentiated by the size of the landmass over which they operate. The Russian landmass

substantially exceeds that of the United States, for example. Agricultural productivity varies; that of Zimbabwe has declined substantially in recent decades. The natural resources of some of these countries are similar. Both Russia and Saudi Arabia hold considerable oil reserves and derive much of their national income from that resource. In theory, at least, all of these countries have access to the same technologies.

The highest life expectancy in this sample of countries is experienced in the United States and the United Kingdom, both of which have an average per capita GDP which are multiples of those observed for the other countries listed. The GDP per capita of the United States is also substantially above that of the United Kingdom.

I will place the economy of imperial Rome alongside later economies in order to identify the level of its achievements and to understand possible reasons for those achievements.

1.4.2 Relationships between numbers

I have selected a variety of measures as being likely to best explain the performance and achievements of the imperial Roman economy and to place it in comparison to later economies. The main parameters that I will examine are set out in the following table:

Table 1.2 List of key parameters to explain economic performance

Grain yield in kilograms per hectare
Size of the largest cities
Processing capacity of the ports
Life expectancy of the population
Silver content of the coinage
Daily reach within the land transport system
Surplus and deficits in the state finance

This approach enables the comparison of disparate economies, even when there are significant institutional differences. There are relationships between these different variables. At its most basic there is a relationship between agricultural productivity and the size of

cities. More productive agriculture produces greater surplus which can be used to support higher levels of urbanisation.

1.4 The experiment

We can view the economic structure that existed in the period from the beginning of the first century AD to the end of the third century AD as an experiment. Over the period from some time before the first century AD to some time after the beginning of the eighteenth century AD, and some time much later than that in parts of the world, there was very limited change in the technology and knowledge used to produce goods and services. The period from the beginning of the first century AD to the end of the eighteenth century AD was, however, a period of significant institutional, cultural and social change in the territories covered by the Roman Empire. The relative performance of the economy of imperial Rome under its institutional and other structures when compared to medieval and later economies, in this experiment, will help identify which of these structures hindered economic performance and which of these structures facilitated economic performance. The question is: How well did the economy of imperial Rome perform in terms of the quantity and quality of goods and services produced relative to the technology available and in comparison with medieval and later economies and which institutions contributed to or hindered economic achievement?

CHAPTER 2: The agricultural sector and why productivity far exceeded medieval levels.

2.1 Introduction

The Roman economy was principally agrarian: most of the economic activity was directed to the production, distribution and consumption of agricultural product and most of the agricultural product used for food was grain. This chapter describes the principal economic characteristics of the production and distribution of grain.¹⁴ The principal economic parameters are quantified. These parameters have a fundamental bearing on an understanding of the wealth and living standards in this economy. Wide ranges of figures have been produced by scholars for many of these parameters and it is close to impossible to form a clear view of the structure of this economy without clarity in this area. A main purpose of this chapter is to provide that clarity.

A principal economic advantage of grain over other foodstuffs is that, with some basic precautions, it does not rot or deteriorate in its unprocessed state in the way that almost all other foodstuffs do. Grain can therefore be transported over great distances in a 'raw' state. The maximum distance between the consumer and the producer of grain is not determined by the speed of transport as is the case with perishable goods. Furthermore, grain can be stored through the year for consumption as needed and the seasonality of its production is not a determinant of consumption patterns. Thus, grain can be consumed at a great distance in space and time from where it was produced and it allows centres of food consumption to exist which are separate from the centres of production. Grain has been, therefore, a major enabling factor in the process of urbanisation and of economic development. In the next chapter I will present a model which identifies the maximum levels of urbanisation that were possible given the levels of agricultural productivity. The model used takes account of the capabilities of the available transport systems and takes account of agricultural production strategies.

There are wide variations in annual grain yields that occur because of changes in the weather. To illustrate this variability: conditions in the United Kingdom in 2012 were such that crop yields fell by between 30% and 50% and much of what was produced was useless for either

¹⁴ In addition to the works cited, I have relied on discussions with Mr. P. Mulcahy, an agronomist practising in the United Kingdom, for much of my understanding of grain and its cultivation. Errors in that understanding are mine. The translations of the Latin texts are my own.

milling or malting.¹⁵ This chapter also illustrates the importance of thinking about the productivity parameters as variables. For example, the incidence of famine is related to socio-economic structures which could not withstand normal levels of agricultural variability, rather than to low levels of agricultural technology.

Finally, the average level of grain productivity that can be achieved in an economy is a partial measure of the level of development of that economy. This chapter therefore also aims to provide a quantification of where this economy sits between Neolithic, medieval and modern economies, and to identify the cultural factors that limited its progress. In short, I develop a parameter which can be used to map the broad structure of economic development over time.

2.2 Seed Productivity

This section explores seed productivity in the ancient Roman world with the intention of positioning that world in the range of economic development and with the intention of explaining the levels of urbanisation, for example, that were achieved. The first step is to quantify the potential range of yields per hectare that could be achieved. A wide range of numbers can be produced from references to yields and sowing rates found in the literary sources. My approach is to put these references in the context of the credibility of the individual writers, who were not equally knowledgeable or reliable. The opinions of some of the ancient writers on some matters are to be completely discounted and on others are relevant only to the extent that they illustrate the extent to which knowledge and practice can deteriorate over time. Elimination of the confusion that results from the Comparisons are drawn with modern yields to identify those areas of economic growth which were within reach of the ancient economy and those areas which could not have been achieved. Potential reasons for the failure to achieve available growth potentials are identified in Chapter 6 Knowledge.

2.2.1 Analysis of Roman grain productivity

Current analyses are based on two main different methodologies. The first approach is to review the numbers contained in the ancient literary sources. A second approach is to use qualitative data (description of husbandry, knowledge of local climate, soil and other relevant variables) to find more modern examples where these variables are similar and the

¹⁵ Discussion with P. Mulcahy.

agricultural outcomes are known and to use those more modern outcomes as representative of what was achieved in the past.¹⁶

Cicero in his speech against Verres when he is attempting to show that Sicily was subject to extortion by Verres in the 70s BC, and explaining variability by reference to divine intervention, says:

*in iugero Leontini agri medimnum fere tritici seritur perpetua atque aequabili satione; ager efficit cum octavo, bene ut agatur; verum ut omnes di adiuvent, cum decumo.*¹⁷

*The longstanding and uniform practice in the territory of the Leontini is to sow about one medimnus of wheat per iugerum. If it grows well it produces a yield of eight to one. But if all the gods apply their help then it produces a yield of ten to one.*¹⁸

Varro in his *Res Rusticae* from the late first century BC adopts a scientific approach, observing variations in experience and attempting to find explanations for that variation:

*seritur fabae modd IIII in iugero, tritici V, hordei VI, farris X, sed non nullis locis paulo amplius aut minus. si enim locus crassus, plus; si macer, minus. quare observabis, quantum in ea regione consuetudo erit serendi, ut tantum, facias, quod tantum valet regio ac genus terrae, ut ex eodem semine aliubi cum decimo redeat, aliubi cum quinto decimo, ut in Etruria locis aliquot. in Italia in Subaritano dicunt etiam cum centesimo redire solitum, in Syria ad Gadara et in Africa ad Byzacium item ex medio nasci centum. Illud quoque multum interest, in rudi terra, an in ea seras, quae quotannis obsita sit, quae vocatur restibilis, an in vervacto quae interdum requierit.*¹⁹

Four modii of beans per iugerum are sown, five of wheat, six of barley and ten of spelt. A little more or less is sown in some places. More is sown on rich land and less on thin. So note the normal sowing practices in each area so that you apply enough to enable the locality and type of soil to achieve its best. The same seed will produce a yield of ten times in one place and fifteen times in another, as in certain places in

¹⁶ Halstead (1987) cautions against too readily comparing yields from ancient and more recent times. Part of the comparative process must include a deep understanding of the agricultural context. It is not sufficient just to compare the numbers. The work by Spurr (1986), referred to later in this section, passes this test.

¹⁷ Cicero, *In Verrem* 2.3.112.

¹⁸ Own translation.

¹⁹ Varro, *Res Rusticae* 1.44.1-3.

*Etruria. I am told that in Sybaris in Italy you usually get a return of as much as a hundred times, and in Gardara in Syria and Byzacium in Africa you also usually get a yield of a hundredfold. It also matters a lot whether you are sowing on land which is untilled, or on land which is used to grow other plants or on land which lies fallow.*²⁰

Varro describes a rich experimental approach for knowledge development:

*bivium nobis enim ad culturam dedit natura, experientiam et imitationem. antiquissimi agricolae temptando pleraque constituerunt, liberi eorum magnam partem imitando. nos utrumque facere debemus, et imitari alios et aliter ut faciamus experientia temptare quaedam, sequentes non aleam, sed rationem aliquam: ut si altius repastinaverimus aut minus quam alii, quod momentum ea res habeat, ut fecerunt ii in sariendo iterum et tertio, et qui insitiones ficulnas ex verno tempore in aestivum contulerunt.*²¹

*In fact nature gave us two ways to develop agriculture, namely experimentation and imitation. The original farmers established many practices by trial and error and their descendants by imitating them, for the most part. We ought to apply a combination of these methods, imitating other people and ways in order to generate experiments to test certain hypotheses, not in a random way but systematically. So we might do the second ploughing more or less deep than others do to see if ploughing depth matters. It is the same approach used by those people who weeded a second and third time and those who deferred the grafting of figs from spring into summer.*²²

Pliny in his *Naturalis Historia* (pre-AD 79) also gives figures for the yields that can be obtained and for sowing rates. He adopts a less scientific approach. He marvels at the miraculous prodigies of nature and is averse to experimentation. He may well be selectively copying Varro, and in any case disagrees with Cicero:

tritico nihil est fertilius – hoc ei natura tribuit quoniam eo maxime alebat hominem – utpote cum e modio, si sit aptum solum quale in Byzacio Africae campo, centeni quinquageni modii reddantur. misit ex eo loco divo Augusto procurator eius ex uno grano – vix credibile dictum – cccc paucis minus germina ...cum centesimo quidem et

²⁰ Own translation.

²¹ Varro, *Res Rusticae* 1.18.7-8.

²² Own translation.

*Leontini Siciliae campi fundunt aliiue et tota Baetica et in primis Aegyptus. fertilissima tritici genera ramosum ac quod centigranium vocant...*²³

*Nothing is more fertile than wheat seeing that from one modius you get one hundred and fifty back, if the soil is suitable as it is on the Byzacium plains of Africa. Nature gave wheat this fertility because she used it to supply most of man's nourishment. Although it is almost impossible to believe, the procurator of that region once sent to the divine Augustus the nearly 400 shoots that had come from one seed. Indeed the Leontini plains of Sicily abound in yields of a hundred fold, as do other places and all of Baetica and especially Egypt. The most fertile species of wheat is the branched type and the so-called hundred grain variety.*²⁴

*...non transferendum est ex frigidis locis semem in calida, neque ex praecocibus in serotina nihilque in contrarium ut praecepere quidam falsa diligentia. serere in iugera temperato solo iustum est tritici aut siliginis modios v, farris aut seminis, quod frumenti genus ita appellamus, x...in denso aut cretos aut uliginoso tritici aut siliginis modius 6, in solute terra et sicca et laeta iv; macrum enim solum, nisi rarum culmum habeat, spicam minutam facit et manem ..emittunt.*²⁵

*Seed which is used in cold climates is not to be used in warm climates and you are not to use early ripening grain in place of late ripening grain. Nothing is to be done which is against normal practice as some people advise us to do out of a misleading earnestness. It is reasonable to sow five modii of wheat or winter wheat to each iugerum and ten modii of spelt or of that seed which we call grain ... in solid, chalky or marshy soil six modii of wheat or spelt should be sown. Indeed this soil produces ears which are small and empty unless the stalks are planted far apart.*²⁶

Columella writing in the late first century AD describes sowing rates and yields as follows and recognises that there are differences in opinions:

iugerum agri pinguis plerumque modios tritici quattuor, mediocris quinque postulat; adarei modios novem, si est laetum solum, si mediocre, decem desiderat. nam quamvis de mensura minus auctoribus convenit, hanc tamen videri commodissimam

²³ Pliny the Elder, *Naturalis Historia* 18.94-5.

²⁴ Own translation.

²⁵ Pliny the Elder, *Naturalis Historia* 18.197-200.

²⁶ Own translation.

*docuit noster usus; quem si quis sequi recusat, utatur praeceptis eorum, qui bene uberem campum in singula iugera tritici, quibus et adori octo modis obserere praecipunt atque hac portione mediocribus agris semina praebenda censent.*²⁷

*A iugerum of rich land demands a full four modii of wheat, middling land demands five; spelt desires nine if the soil is full of joy, and ten if it is unremarkable. For although the writers on this subject suggest a lower amount, our practice indicates to us that this is the most appropriate. Anyone who rejects my amounts can follow the approach of those people who instruct us to sow eight modii of spelt per iugerum on fertile plains and who think that seed should be offered to unremarkable soil in the same proportion.*²⁸

*nam frumenta maiore quidem parte Italiae quando cum quarto responderint, vix meminisse possumus. Cur ergo res infamis est?*²⁹

*Indeed, it is scarcely possible to remember a time when seed produced a four-fold yield in the greater part of Italy. So why is (viticulture) infamous?*³⁰

In addition to the evidence from the literary sources, there is a concentration of evidence contained in papyri containing third-century estate accounts found in Egypt. Rathbone summarises the evidence from the third-century estate of Appianus as indicating sowing rates of 1 *artaba* to the *aroura*. This sowing rate was a norm throughout Egypt. The evidence attests yields ranging between 7 and 16.6 *artabas* per *aroura*.³¹ This documentary evidence is unaffected by the distortions that arise from exaggerated reporting or from the distance between the area being described and the reporter.³²

²⁷ Columella, *de Re Rustica*, 2.9.1-2.

²⁸ Own translation.

²⁹ Columella, *de Re Rustica*, 3.3.4.

³⁰ Own translation.

³¹ Rathbone (2007b) 703-4.

³² These figures can be supplemented by figures taken from *P. Bingen* 111 recto.i. The record of harvest from 3 plots of directly worked (i.e. not leased) land are as follows: 1) 260.1 *artaba* from 18 *arouras*, 2) 220.2 *artabas* from 15 *arouras* and 3) 346 *artabas* from the 'crown land'. Other texts show that the estate held 42.5 *arouras* of crown land but not all of it was necessarily under wheat. Details supplied by D. Rathbone. The first two figures convert to about 1,600 kg per hectare and the third figure produces a similar result if we assume that about half the crown land was set to wheat. These figures can be further supplemented by *P. Colt* 82 which dates from the seventh century A.D. and comes from the Negev in Israel, which is an arid region. This document gives an instance where 40 *modii* of wheat were sown and yielded 270 *modii* of wheat and a second instance where again 40 *modii* of wheat were sown and 288 *modii* were reaped. The yield documented here is about 7 to 1. See Mayerson (1984). If we assume a sowing rate of 4 *modii* per *iugerum* this indicates plots of size 2.5 hectares and a yield of 800 kg per hectare.

White argued that the literary evidence indicates that yields had been in the range of 10:1 to 15:1 in the late first century BC and noted that these yields were consistent with yields seen in the 1950s in Italy where processes and tools were similar, in his view, to those used in the time of the agronomists. The very high yields of 100:1 he attributed to the large number of grains per plant that can be achieved in dry-farming with low density sowing; this high yield per plant is not indicative of high yields per unit of land. The low yield cited by Columella he considered to be that achieved where grain was intersown with vines.³³

Evans regarded the results reported by Cicero and Varro as maxima and concluded that Roman agriculture could not generate a long-term surplus.³⁴

Spurr argued that ranges of outcomes should be stressed rather than averages which understate variability and noted some of the causes for variations: regional differences, seasonal changes, quality of the workforce and the fact that yields are higher on plains than in hills. He took the yields of 10:1 to 15:1 as possibly representative for Etruria.³⁵ The low figure cited by Columella he attributed to an attempt to underplay the economic advantages of cereals relative to vines.³⁶ Spurr gave the literary evidence no more weight than he gave to data drawn from a sample of tenth to twentieth century AD farms where cultivation practices were comparable to those used in Roman times. These data give yields in the range of less than 2:1 and up to 16:1 and reflect variations in farming practices and quality of soil. The lowest figures are from mountain and hill farms and where primitive farming practices such as the use of the spade rather than a plough to till the land were employed.³⁷

Sallares, writing about the Greek world but considering the Roman evidence and potential yields, stressed the importance of an understanding of soil and its management as a determinant of productivity. High concentrations of the essential elements of nitrogen, phosphorus and potassium in the soil will support high yields but, without replenishment of these elements extracted by growing crops, soil becomes exhausted and incapable of

³³ White (1963) 207-12.

³⁴ Evans (1981) 428-42.

³⁵ The emphasis that Spurr (1986) places on variability and uncertainty is essential to the understanding of seed productivity since crop yield is a stochastic variable. With fixed parameters for the independent variables which include seed type, local climate, technology and land management practices the dependent variable of yield fluctuates over time. The reduction in yields in the United Kingdom during 2012, already referred to, illustrates this point.

³⁶ Following Frank (1940) v.141.

³⁷ Spurr (1986) 82-88.

supporting these yields and productivity falls.³⁸ Farmed land is not in itself productive: it is the processes and materials that are applied to it that keep it so or make it so. For example, allowing land to lie fallow enables natural processes to extract nitrogen from the atmosphere and fix it in the soil, while the application of manure can supply elements removed by crops as can the growing of legumes for fodder. Sallares believed that practices of replenishing soils in antiquity were limiting factors as was the low productivity of the seed types then in use. He supported the Columella figure as being consistent with the low economic value placed on cereals relative to vines and the low productivity achieved in medieval Greece and medieval Europe – excluding cases where large amounts of fertiliser were available - but read it as an average rather than a maximum. He made no reference to the other literary sources.

Erdkamp on the contrary regarded the various yields mentioned in the agronomists as credible. He stressed the high degree of variability of results achieved on land of differing quality: yields on first-quality land are about twice those on third-quality land. He read the low figure of 4:1 as either unreliable or representative of what was achieved on poor land and that yields on the better quality land which tended to be cultivated by wealthy landowners were in the range 8:1 to 10:1 but these latter figures were not to be regarded as averages. These higher yields are consistent with what was sometimes achieved in early modern Europe.³⁹

Comparatively little attention has been paid in the review of the literary sources to the sowing rates, which are treated reasonably consistently.⁴⁰ Cicero sets the sowing rate for wheat at one *medimnus* or six *modii*. Varro gives five *modii*, Pliny between 4 and 6 *modii* and Columella says between four and five *modii* depending on the land with up to 8 to 10 being possible on medium land. Sowing rates of 4 to 6 *modii* per *iugerum* equate to between 110 kg/ha and 170 kg/ha and are broadly consistent with modern sowing rates for wheat. The high figure of ten *modii* that Columella gives corresponds to about 280 kg/ha, is substantially above the figures provided by the other sources and is also high by modern standards. Both Pliny and Varro

³⁸ Most crops increase yields pro rata with nitrogen fertiliser but need a supply of the 'basic' fertilisers of phosphate and potash. The skill comes in determining how much to apply, before you induce lodging – which is to say cause the crop to fall over. Nitrogen can be supplied by first growing leguminous plants which absorb nitrogen into their roots. Leaving these roots in the ground gives a supply of nitrogen for wheat subsequently planted. Manure is another source of nitrogen. See also Sallares (1991) 381-389 for discussion of fertiliser. The essential elements of the chemistry are that chlorophyll, which is used by plants to convert water and carbon dioxide into sugars, requires nitrogen. Wheat cannot access the abundant nitrogen in the atmosphere and needs to take this element from the soil. The supply of nitrogen to the soil can be enhanced through the use of natural fertilisers such as manure or chemical fertilisers.

³⁹ Erdkamp (2005) 34-46.

⁴⁰ Evans (1981) 434 notes the similarity of the reported sowing rates.

suggest sowing rates for spelt (and emmer in the case of Pliny) of 280 kg/ha.⁴¹ Sowing rates in Egypt were about 1 *artaba* per *aroura* or about 112 kg/ha.

Table 2.1 shows the yields in kg per hectare claimed by the various sources.⁴²

Table 2.1 Summary of grain yields

1: Source	2: Region for which estimate given	3: Time	4: Sowing Rate <i>modii/iugerum</i>	5: Yield	6: Yield kg/ha	7: Commentary on estimate
Varro (1)	Etruria	Late 1 st c BC	5	10:1 to 15:1	1,400 to 2,100	Illustrative?
Varro (2)	Sybaris, Gadara, Byzacium	Late 1 st c BC	5	100:1	14,000	Illustrative
Cicero	Leontini in Sicily	70s BC	6	8:1 to 10:1	1,300 to 1,700	Favourable to exceptional conditions
Columella	Greater part of Italy in living memory	Late 1 st c AD	4 to 5	4:1	450 to 550	Presented as a maximum
Pliny	Byzacian plain, Leontine plain, other parts of Sicily, Baetica, Egypt	Late 1 st c AD	4 to 6	100:1 to 150:1	11,200 to 25,200	Presented as results for areas with suitable soil
Heroninos Archive	Egypt	3 rd c AD	4	7:1 to 17:1	800 to 1,800	Documentary

⁴¹ The high sowing rates for spelt are puzzling.

⁴² I have used the following conversion factors in moving from ancient measures to modern equivalents: 1 *artaba* of Egyptian wheat equates to 30.28 kg/38.78 litres of wheat, one *modius* of wheat to 6.8 kg/8.62 litres of wheat. An *aroura* is equivalent to 0.276 hectares and a *iugerum* is equivalent to 0.251 hectares.

The high Varro (1) figure is made with particular reference to Etruria and is intended to represent the very favourable outcomes achieved in some localities. The lower figure is intended to be a more general average. The Varro (2) figures can be dismissed as not representative of any plausible economic yield. They reach the highest that can currently be exceptionally achieved with the best of modern technology. They may relate to observations of single exceptional plants rather than what was achieved over any appreciable area, and are intended to illustrate the high level of variability in yields and the exceptional fertility of the soil of Etruria. They are also presented as hearsay figures rather than the author's direct experience. The results derived from Pliny are beyond the genetic potential of the seed material available in his time and at the higher end of his implied range exceed the maximum yield that agronomists think that land will ever achieve.⁴³ Pliny is not describing a reality and his evidence in this connection must be dismissed as completely unreliable.

The figures provided by Cicero relate to fertile soil in Sicily and are described as being for favourable or better climate conditions. Long term averages – which will include unfavourable outcomes – were therefore perhaps 1,200 kg/ha.

Columella reports figures which are intended to be maxima and which are intended to apply over a greater part of Italy over a considerable period of time. However, Columella differs from Varro in his views on sowing rates for poor versus fertile soil. For example Varro correctly claims that higher sowing rates should be used on fertile ground. Columella takes the contrary, and incorrect, view that a lower sowing rate should be used. Pliny agrees with Varro and correctly identifies the reason why there is a higher sowing rate on fertile ground. Columella also completely fails to understand the essential importance of leguminous plants as sources of nitrogen.⁴⁴ Columella is either setting down his own inaccurate views or is describing agricultural practices which had degenerated since the time of Varro. The results that he suggests are comparable to those achieved in Neolithic times.⁴⁵ As is frequently noted

⁴³ Kucharik and Ramankutty (2005) 2 indicate 24 tonnes per hectare as the ceiling.

⁴⁴ At 2.10.7 he expresses the view that planting beans is more effective than fallowing. In other words he fails to understand that it is precisely the opposite that is true; beans fix atmospheric nitrogen into the soil.

⁴⁵ See Table 2.2.

the reference to this figure is contained within an argument for the economic benefit of viticulture and may have been constructed for rhetorical reasons. Rather than completely discard this figure, I will use it as an indication of results achieved where agricultural knowledge and practices are poor.

The figures for Egypt partly reflect the fact that the yearly inundation of the Nile served to replenish the soil. Silt carried with it elements removed by the growth of the previous year's crops and to an extent did naturally what required human intervention elsewhere. It is possible that the relative ease with which grain could be accumulated and distributed along the Nile and across the Mediterranean through the port of Alexandria incentivised the development of sound land and crop management practices.

These figures may be read as indicating yields of 1,200 kg/ha to 1,800 kg/ha depending on soil type and local conditions and where farming practices were reasonable competent. The lower end of the range is representative of average experiences in many parts of Italy, for example, and the higher end of the range is representative of average experiences in Egypt and Etruria. Figures outside this range can be expected to occur because of normal variations in weather conditions. Yield of below 400 kg/ha will occur where farming practices are poor. With the exception of the Varro (2) and the Pliny figures, all of these yields lie within the range identified by Spurr. Taken together these figures imply high degrees of geographical variation and possibly more than implied by the variations in soil type, land position and local climate.

In what follows I will use a figure of 400 kg per hectare as what was achieved under stressed conditions – where soil was depleted, where farming practices were poor, for example - and I will use a figure of 1,200 to 1,800 kg per hectare as what was achievable where land management practices and natural conditions were reasonable, with significant variations outside this range possible because of the weather.

2.2.2 Comparisons to other economies

There is a correlation between the level of agricultural productivity and overall economic performance, and population. Higher agricultural productivity leads to higher agricultural surplus which allows, and indeed requires, greater urbanisation, with consequent opening up of opportunities for trade, specialisation and development of the non-agricultural sector. Grain productivity is an important measure of economic performance not only because it has these consequences but also because it reflects other activities and structures in the economy.

So, for example, the development of the scientific experimental approach led to improvements in grain productivity. Institutional structures such as the forms of land ownership, and the availability of investment opportunities for savings can also affect grain productivity. These aspects will be explored later in this chapter and in subsequent chapters.

Table 2.2 on the next page compares my estimate of 1,500 kg per hectare average performance for imperial Rome, which is the middle of the range of figures of 1,200 kg to 1,800 kg referred to above, with earlier and later performance, for some territories covered by the Roman Empire.⁴⁶ This economic map shows the economic performance during imperial Rome as being twice that of Neolithic times. After our period, performance dropped to less than half.⁴⁷ It was not until the middle of the eighteenth century that performance recovered. Estimates for modern times are significantly above those of eighteenth century Europe and of imperial Rome. Much of that later improvement follows on from technological change.

⁴⁶ The estimate shown for Neolithic times in the table is based on Gregg (1988) 70-5 (but adjusted as described further below) who gives a range of mean yields for a variety of wheat types. The low end of the range is 600 kg per hectare for spring einkorn and the high end of the range is 1,000 kg per hectare for winter emmer-spelt. Gregg placed most emphasis on climate as the differentiator of performance in Neolithic times and based her yield figures on data drawn from Germany in the period 1850 to 1905, this period being chosen because climate conditions were similar. The seed types used in both reference points are the same and she implicitly considered that any differences in farming practices were irrelevant to conclusions on yield. I believe that the implied mid-range figure of 800 kg per hectare is too high, since I believe that land and crop management practices in Germany in the period 1850 to 1905 were very likely to have been superior to those used in Neolithic times. I have therefore judgmentally selected a figure of 500 kg per hectare for Neolithic times. The figures for medieval experience and for Western Europe are taken from Slicher van Bath (1963) and Broadberry, Campbell, Overton, Leeuwen and Apostolides (2009). Slicher van Bath (1963) compiled seed yield figures for various crops. Broadberry et al (2009) 30 Table 2 shows English wheat yields for 1250 to 1499 averaging at 7.36 bushels per acre, net of seed, which I convert to 620 kg per hectare gross of seed, converting 1 bushel to 27 kg, 1 acre to 0.4 hectares and assuming that 20% of the gross output is used for the following year's seed. The figures for 1700 to 1799 average at 15.53 bushels per acre, net of seed, which I convert to 1,310 kg per hectare, gross of seed. The estimates for modern times relate to the period 1993-1995 and are taken from <http://www.fao.org/docrep/006/y4011e/y4011e04.htm>. Retrieved 30 October 2015.

⁴⁷ Wickham (2004) surveying the period AD 400 to AD 800 concludes that it was a period of economic decline. Fiscal structures became much simpler, disappearing in the Roman-Germanic kingdoms and surviving in the East only in fragmented forms. There was a decline in elite wealth and a decrease in exchange. This reduction in fiscal and elite resources led to a great simplification in material culture. The period was also one of increased instability.

Table 2.2 The economic map**Economic progress measured by grain productivity**

	Amounts in kg per hectare
Estimates for Neolithic experience	500
Roman experience as per documentary/literary sources	1,500
Estimates for medieval experience	600
Estimates for Western Europe eighteenth century	1,300
Modern comparators	2,000 to 7,500
Italy	3,400
Spain	2,000
United Kingdom	7,500
Egypt	5,300

As indicated in Table 2.2 the level of agricultural production in medieval Europe is not far above that achieved in Neolithic times. This comparison suggests that medieval land and crop management practices were primitive. Medieval yields were not very different to what could be achieved with no human intervention beyond sowing and harvesting. Gregg notes that at this level of productivity the amount of nitrogen being extracted from the soil by the growing

crops is about the same as that naturally replaced by rain, dust, birds, weeds and seed.⁴⁸ The fact that these results are being observed indicates that fertilisation procedures were primitive.⁴⁹ Weeding is important to maximising the yield produced on cultivated crops, reducing the nutrients, light and water absorbed by competing plants. Again the yields achieved suggest that limited weeding was carried out. This work is laborious and unlikely to be undertaken if the free agent who might carry out this work will not benefit from the additional yield or if those who control the labour of others do not benefit from the increased yield. It does not matter much at this level of productivity whether the land is allowed to lie fallow or not, and the available statistics suggest that such land practices had no impact on yields.⁵⁰

Comparison of yields achieved in the time of imperial Rome with those later achieved in medieval times indicates a severe degradation in land and crop management practices, which may be related to structural disincentives.⁵¹ It was not until the middle of the eighteenth century in Western Europe that yields recovered to the imperial Roman level.⁵² The period from imperial Rome through medieval times and into the middle of the eighteenth century is not a period of significant technological change. It is, however, a period of profound social and cultural change. Figures 2.1 and 2.2 shows grain yields for various European countries and illustrate a process of stagnation followed by continuing incremental growth. The first graph shows the development from 1250 to 1800 in England.⁵³ The second graph shows a

⁴⁸ Gregg (1988) 70-5.

⁴⁹ The effects that fertiliser can have on yields is illustrated by Humphries and Biffen (1907) 7 which shows the yields for unmanured land as being about one third that of manured land. The higher yields are, however, offset by deterioration in baking quality.

⁵⁰ The compilation of wheat yields produced by Slicher van Bath (1963) does not show descriptions of soil type or farming practice before the seventeenth century, with only one set of exceptions to this. The exceptions relate to yields achieved in Provence in 1338. These yields are shown for 26 landholdings divided into the following farming practice categories; a) two-course system with fallow, b) two-course system without fallow (or only partly fallow) and c) three course systems (sometimes with fallow during two out of the three years). There is no clear indication from these figures for Provence in the fourteenth century that any of these systems was relatively advantageous. I suspect soil types and farming practices were not recorded before the seventeenth century because they very largely did not matter. There is a greater level of differentiation in yields starting in the seventeenth century and more commentary on soil type.

⁵¹ Most of the elements required by the crop are provided for free, as is the energy. Carbon, oxygen and hydrogen and energy are provided by water, the air and sunlight which are abundant relative to the requirements of the crops growing to uncultivated yields. The element which is not free, beyond the primitive yield, is nitrogen because wheat cannot absorb nitrogen from the atmosphere. The fact that primitive yields are not substantially exceeded in medieval times suggests that limited work was carried out to maximise crop yields.

⁵² The recovery did not happen in Russia until much later.

⁵³ The first graph is based on figures from Broadberry et al (2009) 30 Table 2 converted to kg per hectare, gross of seed.

wider European comparison.⁵⁴ I have used the first graph as a measure of how agricultural productivity varied over the period 1250 to 1800 and as a basis for comparing that productivity to what was achieved in the time of the Roman Empire. I have used the second graph to test whether the conclusions drawn from the English data apply more widely. The yields shown in both graphs are to be compared with a Roman imperial experience of 1,500 kg per hectare. The first graph shows that agricultural productivity in England did not reach the level which had been achieved in the time of imperial Rome until about 1750. The second graph shows that grain productivity in England was the highest in Europe from the beginning of the fifteenth century and that no country in Europe equalled the performance of imperial Rome until about 1750.⁵⁵

⁵⁴ The figures in this graph are taken from Slicher van Bath (1963) wheat yields. Just over 1,500 wheat yield figures with associated date and territory were typed in and then sorted by year and territory to generate this graph. The yields given have been averaged into 50 year groupings to dampen down the effects of stochastic variation and to make the underlying changes in the mean clearer. The yields were then multiplied by 140 to convert from wheat seed yields to yields per hectare.

⁵⁵ The general level of the grain yields for England in both graphs are the same but there are differences in the shapes of the two graphs. I attribute these differences to possible differences in the samples on which the graphs are based and to the high level of variability associated with grain yields. Long histories of grain yields for the same locations were used for the first graph. A large collection of isolated observations produces the second graph.

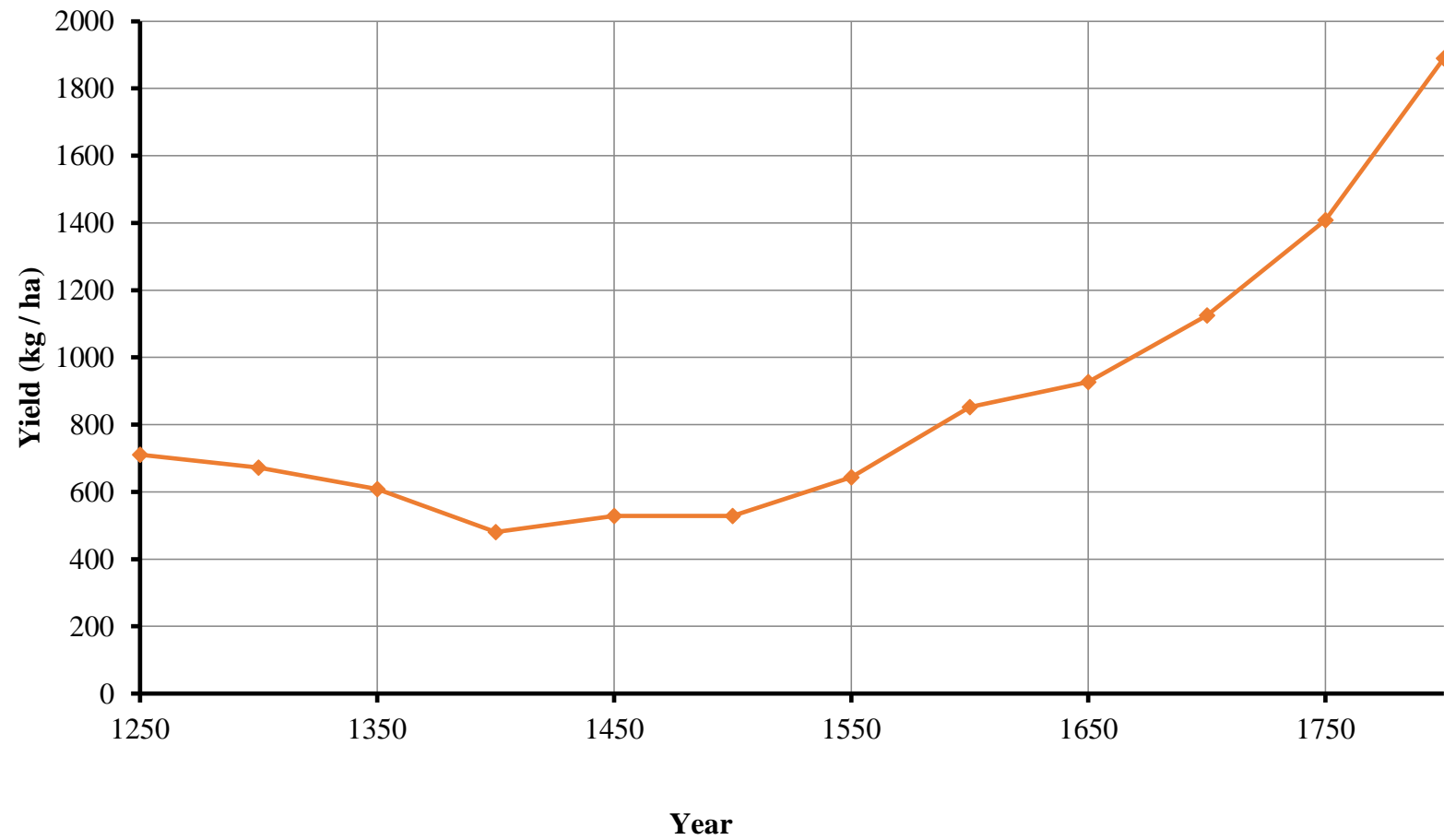
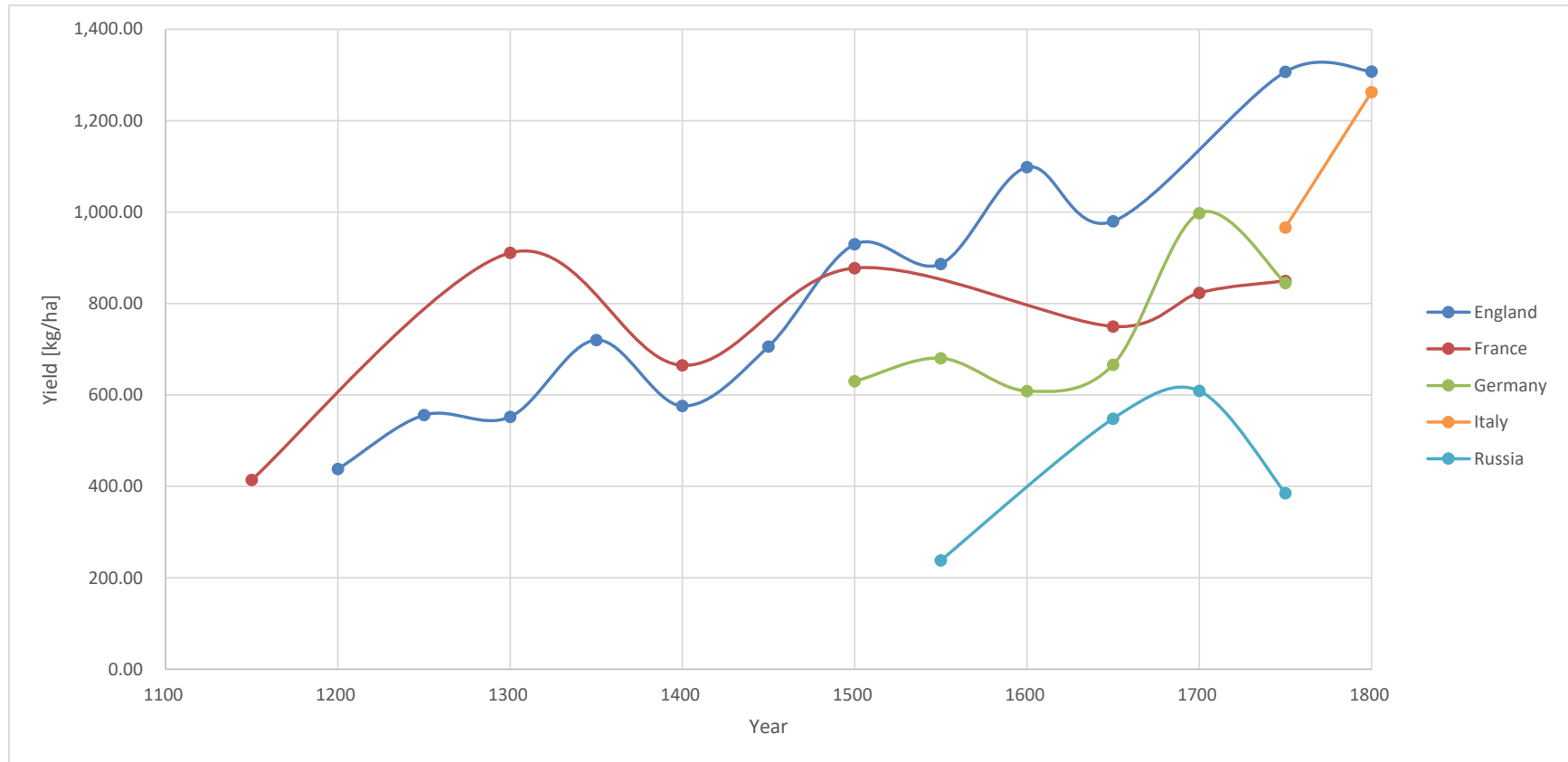
Figure 2.1 History of English grain yields

Figure 2.2 History of various European grain yields

The upturn in the English grain yields happens around about the time of the dissolution of the monasteries, which took place in the first third of the sixteenth century. The change in land ownership that followed may have changed the incentives to extract more value from the land; this is discussed later. More general cultural shifts may have had an impact on work practices.⁵⁶ Sallares also indicates that the Agricultural Revolution that took place in early modern England followed from a change in land use. Land was used for a number of years to grow cereal and then as the land became less fertile with the extraction of chemical elements it was put to use as pasture which was then fertilized by animal manure. These changes in the use of land doubled food production in England. This approach requires that there be an integrated approach to the rearing of animals and the growing of crops.⁵⁷ By the eighteenth century yields in England were not only substantially higher than in medieval times but also more variable between types of land.⁵⁸ I attribute this higher variability to improvements in land and crop management practices. When these were at the relatively primitive level of medieval times poor and rich land yielded much the same. As practices improved the yield on poor land reached a limit while the yield on rich land increased beyond that to a new maximum.

In order to fully understand the changes that occurred in agricultural productivity in the period from the time of imperial Rome to the middle of the eighteenth century we need to understand the process of surplus production. We also need to understand, and put some numbers around, the process of surplus production in order to be able to link levels of agricultural productivity and levels of urbanisation. The definition of surplus that I will use is, for any production unit:

Grain surplus = Grain produced less grain consumed on the production unit Formula 2.1

I will, as before, limit myself to wheat. In grain consumed on the agricultural unit I include that consumed by those who live or work on the unit and that consumed as seed. This surplus is not profit. Tenants need to deduct rents paid and both owners and tenants need to deduct

⁵⁶ A popular Catholic magazine in March 2015 characterised the culture change that took place in England at the time of the Reformation as a movement from a world which "...emphasized ascetism, renunciation of sex and wealth and pleasure, suspicion of intellectuals, corporate rather than private prayer. But now there was manifested a widespread desire for a different form of Christianity, less ascetic, more mundane, more intellectual, more individualistic, more secular, but not less religious ..."

⁵⁷ Sallares (1991) 384.

⁵⁸ Based on a review of yield figures for wheat in Slicher van Bath (1963).

other costs of production. It does, however, provide the link between the agricultural unit and the urban unit because what is not consumed on the agricultural unit is consumed in the town.

2.3 Consumption

Although there is some debate about diet, which clearly varied over time and geography and by social grouping, a figure of 300 kg per capita is a reasonable measure of average annual wheat intake. In times of increased prosperity there was significant consumption of meat and the wheat equivalent of total food intake exceeded 300 kg per capita, as discussed below.

The literary sources provide some detail on the amounts of wheat consumed during the Roman Republic. Cato recommended the allocation of just over 350 kg per annum to those set to work in the fields and about 250 kg to those engaged in less heavy work. In the following passages the principal components of the diet are recorded. The best quality produce is reserved for the customer. Food that could be obtained by foraging or from gardens, berries for example, are not recorded because Cato is thinking of what he can sell. Some of these supplements are essential components of a diet.

familiae cibaria. qui opus facient per hiemem tritici modios IIII, per aestatem modios IIII S, vilico, vilicae epistatae, opilioni modios III, conpeditis per hiemen panis P. IIII, ubi vineam fodere coeperint, panis P. V, usque adeo dum ficos esse coeperint, deinde ad P. IIII redito.

vinum familiae... summa vini in homines singulos inter annum Q. VII. conpeditis, uti quidquid operis facient, pro portione addito; eos non est nimium in annos singulos vini Q. X ebibere.

pulmentarium familiae. oleae caducae quam plurimum condito. Postae oleas tempestivas, unde minimum olei fieri poterit, eas condito, parcito, uti quam diutissime durent. Ubi oleae comesae erunt, hallecem et acetum dato. oleum dato in menses uni cuique S.I. salis uni cuique in anno modium satis est.⁵⁹

Food rations for the household: For those who work in the fields, in the winter four modii of wheat and in the summer four and a half modii of wheat. Allocate three modii to the bailiff, to his wife, to the overseer and to the shepherd. Those in chains are to get 4 pounds of bread during the winter and 5 pounds when they start to dig the

⁵⁹ Cato, *de Agri Cultura* 56-58.

vines and continuing at this rate until the figs begin to ripen after which the allowance drops back to 4 pounds of bread.

Wine for the household: ... the total allowance of wine per individual is seven quadrantals. The amount allocated to those in chains is proportional to the type of work they do, it is not excessive if they each drink ten quadrantals of wine a year.

Relish for the household: Store as much as you can of the windfall olives. After that store the ripe olives that will not produce a lot of oil. Issue these sparingly so that they last as long as possible. When the olives have been eaten, issue fish sauce and vinegar. Issue one pint of oil to each individual per month. A modius of salt per person per annum is enough.⁶⁰

The wheat issued contains all of the necessary calories, more protein than needed and significant amounts of some of the needed minerals and vitamins but is short of vitamins A and C, is low in fat and very low in calcium. The amount expended on wheat, which could be sold and has economic value, is clearly being kept to the minimum necessary for the functioning of the workers and there is no attempt being made to provide a varied or interesting diet.

Polybius quantifies the wheat allowance for soldiers in the Republic at 300 kg.

The allowance of corn to a foot-soldier is about two thirds of an Attic medimnus a month, a cavalry-soldier receives seven medimni of barley and two of wheat. Of the allies the infantry receive the same, the cavalry one and one-third medimnus of wheat and five of barley, these rations being a free gift of the allies; ...⁶¹

No mention is made of the other foodstuffs that the soldier consumed although to function he needed a more varied diet than one based just on wheat.

Davies argues that the average wheat consumption for the Roman soldier of the Principate was one-third of a tonne per annum and also that this diet was substantially supplemented by meat, fish, fruit and vegetables and by wine. He also argues that this diet was similar to that consumed by the civilian population.⁶² Rickman, considering Cato's evidence, uses a figure

⁶⁰ Own translation.

⁶¹ Polybius, 6.39. 13-15. The barley is for horses and is more or less the same as current feeding to a modern horse of weight of 500 kg; the additional wheat allowance for cavalrymen is presumably not for consumption. Loeb translation.

⁶² Davies (1971) 122-141.

of 280 kg per annum as an average consumption rate of wheat and takes the view that grain was the main component of the diet of the majority of the people. Garnsey uses 200 kg as a minimum per capita consumption rate for Rome.⁶³ Allen constructed a diet for the manual labourer which comprised 170 kg of wheat supplemented by 20 kg of beans, 5 kg of meat and 5 litres of olive oil on the basis that this was sufficient to provide the minimum calorie and nutrients needed and was affordable given prices as set out in the Diocletian edict and likely wage levels.⁶⁴

These figures help contextualise the free wheat distribution at Rome. The amount allocated to the citizen at Rome through the wheat distribution was about 400 kg of wheat but this was intended to provide for the man and his dependents, such as his wife and minor children.⁶⁵ This low level of support implies either that the citizen augmented the state support with privately acquired grain or other foodstuffs or that he and his dependents lived at barely or below subsistence levels. In any case it left sufficient incentive to work where possible and favoured the unmarried man with no dependents.

At the different stages of its development the economy supported a population which either very largely depended on wheat as its main food, as Rickman proposes, or had a more varied diet which included meat, as Allen suggests. There is evidence of changes in meat consumption patterns. Jongman collated data which suggests that meat consumption increased in Italy from the third century BC, declined in the second century AD with a subsequent recovery. In the provinces consumption dramatically increased in the mid second century BC and declined in the mid second century AD with no subsequent recovery.⁶⁶

The average amount of wheat consumed per capita during the Roman Republic probably did not vary much from that consumed by soldiers, farm workers and slaves since these are representatives of a large part of the population. It may have been somewhat less than that amount, but not much less, where the rich replaced a part of the calorific and nutritional elements of wheat with the components of a more varied diet. In what follows I will use assume a consumption figure of 300 kg per capita, which is broadly what is mentioned in the sources and which, because it provides just the right amount of much of the basic nutritional and calorific needs of the human animal, when supplemented by fruit and vegetables,

⁶³ Rickman (1980) 10; Garnsey (1983) 118.

⁶⁴ Allen (2009) 337-342. Kron (2008) 83 indicates a diet which had a calorific value of 6,000 which I consider unreliable.

⁶⁵ Sallust, *Historiae* 3.48.

⁶⁶ Jongman (2007a) 191-5.

amounts almost to a biological constant for a diet in which wheat is a significant component. The rearing of cattle or other animals to provide meat requires the production and consumption of fodder or the conversion of arable land into pasture. Broadly speaking 1 kg of meat requires the input of 10 kg of wheat equivalent. Increased meat consumption represents an increase in agricultural activity beyond that represented by the 300 kg figure.

2.4 Labour Structures

This section deals with labour productivity and pays particular attention to the labour structures which were employed in grain production. I use the simplifying concept of grain surplus, the excess of the grain produced on an agricultural unit over the grain consumed on that unit. I stress the importance of understanding the production strategies which were available to the controllers of production, and which they chose to or were constrained to follow. These strategies are an important determinant in total available grain surplus.

White set out a short and comprehensive costing for the labour needed to produce grain using the processes employed in Roman agriculture and based his figures on those provided by Columella. He adopted the same comparative approach as for his work on grain yields and placed the Roman figures alongside the equivalent figures for similar processes as found in sources from the sixteenth century up to modern times. Broadly these figures imply that 50 hectares set to grain requires a workforce of about 8 men, making allowance for down-time due to illness, poor weather and periods when there was limited work for the men to do.

White also positioned these figures as being the labour requirements for large estates which were focused on producing for profit and which employed diversified crop portfolios so that labour was fully utilized, avoiding a problem of monoculture which places periodic demands on labour.⁶⁷

Spurr also costed labour for arable estates using slave labour but unlike White, who considered grain in isolation, he explicitly considered a diversified strategy and evaluated the labour demands of each of the constituent parts. He based his work on an agricultural unit of 50 hectares described by Columella and detailed in the table below and which, according to Columella, required a work force of 8 men. The structure is to be considered an agricultural unit rather than an estate since no mention is made of supervisory, administrative or other overheads.

⁶⁷ White (1965) 102-7. I calculate the figure of 8 men for 50 hectares using White's figure of 9.5 to 10.5 man days per iugerum, converting 1 iugerum to 0.251 hectares and assuming that there are 250 workdays in a year.

Table 2.3 Example structure of agricultural unit

Crop type	Hectares	Man days required
Cereal	12.5	700
Legumes	12.5	500
Three month crops	7.5	300
Dry meadow	7.5	300
Fallow	10.0	0
Total	50.0	1,800

These figures are consistent with the workforce of 8 men as indicated by Columella if allowance is made for downtime. Spurr considered that these agricultural units were so structured that there was no need for hired labour, that is the varying demands on labour were managed by the diversified agricultural structure and the men needed on different parts of the unit at different times.⁶⁸ The unit described above is well structured to provide natural fertilisation through fallow land and through crop rotation. To the extent that Columella is describing units which existed then, agricultural practice reflects a deeper knowledge of fertilisation processes than he is able to articulate.

Erdkamp also evaluated the labour production costs for wheat but considered that Spurr's figures were low because of a failure to allow for supervision costs, upkeep of equipment and other indirect labour costs. He also made the methodological objection that estimating labour requirements by counting the hours needed to carry out a defined set of tasks and then adding those hours up is likely to be optimistic; the worker is not 100% productive all the time. With the exception of the supervisory costs which have to be allowed for somewhere, these objections are too harsh since Spurr made implicit allowance for costs of maintenance and other activities and also because the figures produced by Spurr agree with Columella's

⁶⁸ Spurr (1986) 133-6. Note Rathbone (1981) argues that because of seasonal peaks in demand for labour, as for example at harvest, there was periodic need to supplement slave labour with hired labour and against the possibility of pure slave production.

overall figure of 8 men as being the labour force needed for 50 hectares and must therefore reflect ancillary tasks and downtime.⁶⁹

Erdkamp advanced the debate by using labour productivity figures to articulate some of the main characteristics of the production structures which were used in the cultivation of grain, and used this approach to emphasise the importance of the labour: land ratio. He considered three principal production structures: firstly, a commercial estate of 50 hectares manned by slaves, secondly a one-family household, including hired labourers, on about 14 hectares of fertile soil, and thirdly a peasant family on about 2 hectares of poor land. Of these structures the large estate and the large family farm were capable of producing marketable surplus. The very small farms were capable of producing surpluses only in the most favourable of conditions and were little able to withstand adverse conditions which, because of the labour: land ratio and the poor quality of the soil, were frequent, and at the best of times they could only produce a meagre living.

The conclusions to which Erdkamp comes are enriched by the archaeological evidence which presents a wide range of economic structures and identifies chronological developments in those structures. Rathbone presented a survey of the variability of the individual economic structures that populated the landscape of Roman Italy drawing on literary and archaeological evidence.⁷⁰

Movements upwards in social status required the acquisition of land; acquisition of land beyond the minimum requirements were also socially advantageous. Land transfer under the ownership structures which prevailed in the time of imperial Rome was readily possible. Landholdings could be purchased by new entrants to this social structure or increased in size through payment of monies acquired through a number of routes, including from the payment of proceeds of conquest, army pay and discharge bounties, profits from trade and profits from agriculture.⁷¹ Profits could be generated by renting land and then used to buy land. There

⁶⁹ Erdkamp (2002) 46-54.

⁷⁰ Rathbone (2008).

⁷¹ Some new entrants to the imperial Roman land-owning structure - whether colonists or veterans being allotted land - were given sufficient land to place them in the fifth *classis*. Rathbone (2008) also noted the dynamic nature of the ownership structures with some families prospering and increasing wealth and others falling out of the structure. The status which was linked with greater ownership of land provided a motivation to accumulate larger landholdings through acquisition from other owners and was an incentive to use land productively and to save. Attempts by those receiving land entitlements to cash in and sell to other landowners were occasionally restricted by structuring the land transfer as leasehold with minimum rent or with an outright ban on sales for at least limited periods, that is external restrictions were placed on exits from this property owning structure. The archaeological evidence shows changes in use suggesting migration through the different categories as economic units failed, prospered and were consolidated.

were no upper or lower limit restrictions on the size of land packages which could be acquired. For example, a land holder with a modest holding of 13 hectares, say, managing his land well might in time accumulate sufficient surplus to acquire a neighbour's plot or sell his current holding and together with his savings acquire a larger holding in the same district or elsewhere.⁷² There is, therefore, a connection in this system between holding and social status and behind that a connection between good land management and social status. These landholding and social structures encourage higher land productivity. The manorial land holding system, which prevailed in medieval times, by contrast provides very limited such encouragement. Under the manorial system a single landholding of perhaps hundreds of hectares is held by a lord of the manor. Part of that land he works for his own benefit. Part of that land is worked by peasants for their own benefit, subject to provisions of specified services to that lord and the remainder is similarly worked but for payment of money rents.⁷³ Accretion of landholdings by the lord of the manor was difficult.⁷⁴ It was difficult to gradually add small plots and any acquisition must be a large transaction, subject to overlord approval. The acquisition of land by peasants was also difficult. The practical difficulty of investing in land, under the manorial system, acted as a disincentive to saving. The available investment asset classes were severely restricted.⁷⁵ There is, therefore, a limitation on the incentive to understand and improve land and crop management practices.⁷⁶ Beyond a certain point a good peasant farmer gets no reward for his additional skills and diligence.

The landholding system under imperial Rome was highly flexible. Holdings of about 2 hectares were of the size allocated to veterans and in any case were about the minimum that a man could own and still be considered a member of the property owning classes. The ability to extract value was dependent on the owner's knowledge of best farming practices.⁷⁷ Soldiers who came from farming backgrounds could be expected to have such knowledge but

⁷² '... the Graeco-Roman world was essentially and precisely one of private ownership, whether of a few acres or of the enormous domains of Roman senators and emperors, a world of private trade, private manufacture.' Finley (1999) 29.

⁷³ See Barsby and Barsby (1996) and Fifoot (1932).

⁷⁴ In some cases the acquisition of land under medieval structures was impossible. Villeins, for example, were not allowed to buy on their own behalf.

⁷⁵ The manorial land holding system is relatively inflexible. By contrast the landholding system which applied in the time of imperial Rome was flexible. The emergence of the villa system as the dominant landholding arrangement may have led to restrictions in this asset class.

⁷⁶ In addition, there were restrictions on labour movements limiting the freedom of peasants to leave and the ability of lords of the manor to force them to leave.

⁷⁷ At the very basic level not much was required beyond a combination of what was widely known and hard work.

undoubtedly there was variability. At the low level of productivity reported by Columella these holdings could not support a family.

The small size of these landholdings created volatility of returns which made them unsuitable sources of any stable surplus on which consumers or traders could rely. The variations in agricultural yields that follow from weather variations could push such a holding with a small number of dependents from producing a small surplus to being insufficient to support the owner and his family. Illness or excessive numbers of dependents pushed the economic unit further below the level at which it could sustain itself. The consumption of grain needed for next year's seeding in times of poor harvest carried the effects of adverse conditions of one year into the next year. To this must be added the tendency, as Spurr has noted, of the smallholder to exacerbate his condition by bad farming practices, such as failing to select the best seeds from this year's crop for use in generating the next year's returns. Nevertheless, with a small number of dependents, an able bodied man on good land with a reasonable knowledge of farming practices might be able to sustain himself from one year to the next growing a mixture of grain and other arable crops, combined with a few chickens, a pig and foraging on common land.⁷⁸

For larger holdings the mathematics becomes very different and the available strategies and outcomes more varied. With an increase in the size of the landholdings the size of the agricultural output increases without an immediate need to increase the labour force; labour on the small holdings is very largely unproductive.⁷⁹ The increase in unit holdings therefore enables the owner to increase his consumption above the bare minimum and it also reduces the risk of ruin. With a larger holding surplus was available in almost all years and the main variability was the size of the surplus rather than whether the economic unit was viable.

The controller of production had first call on the economic output, after any rent, and determined the strategy to be applied and use of surplus. The archaeological evidence indicates some variety in these strategies for farms of sizes between 5 and 15 hectares but it shows a general tendency to follow diversified approaches. The remains of these holdings have traces of wine and olive presses and of threshing floors and at the higher end of the range goats, pigs and chickens are kept for consumption and one or more draught animals are

⁷⁸ For a unit of this size to work with a strategy of 50% wheat and 50% legumes and fallow then the wheat yield need to support a family of 4 people is about 1,200 kg/ha.

⁷⁹ One man could tend a plot of between about 5 and 6 hectares in an efficient structure; allocating that man and his dependents to a plot of 2 hectares produces a heavily populated landscape of unproductive labour.

used.⁸⁰ Surplus is expended on the decoration of buildings and some luxury goods. Some of these units support families who adopt strategies which enable them to live well off the produce of their land, supplying most of their needs but with residual surpluses that were used for participating in local economies.⁸¹ The amount of land set to grain or used for other arable purposes or for vines and olives is the outcome of a decision process which reflects the preferences and capabilities of the controller of production, the existing practices on his farm and the availability of markets which could be used for the disposal of expected surpluses.⁸² A farm in this category with a third of its land set to grain and with a household of around 6 individuals might produce a grain surplus available for sale of between 0 and 5 tonnes in an average year.

Landscapes constructed from these units were capable of supplying grain but the amount of grain depended on the controllers of production who generally had first call on surplus and its use. These landscapes were also less densely populated and more prosperous than those constructed from small holdings and are interconnected by the need to dispose of surplus.

Some larger holdings were designed purely for the production of marketable surplus. These units have buildings which are generally much less ornamented and which are purely functional being used for the housing of the manager and the slaves who worked the land, for the housing of elements of the production process such as mills and for storage. In some cases they included living quarters for visiting members of the owner's family. These pure production units ranged in size to around 50 hectares and beyond. Because they were designed to produce surplus for external consumption these units used labour most efficiently; labour on these farms was intended to be as productive as possible and the resources of the farm were not to be consumed by the leisure of those on it. Units of this size can have a large focus on one commodity, although that focus may change over time, for example being largely directed towards producing pigs at one stage and later being largely directed towards producing grain. An agricultural unit in this category with about half of the land, say 25 hectares, set to grain was capable of producing an average grain surplus of

⁸⁰ The absence of remains of goats, pigs and chicken on the smaller sites may be an accident of preservation. The smaller sites almost certainly contained some such animals.

⁸¹ In some cases families, especially at the upper end of the social spectrum may have lived in nearby towns rather than on the farm. The extent of the surplus produced also depends on the relative weight given by the controller of production to leisure and to surplus. Where leisure is more valued less surplus is produced.

⁸² Equally the controller of production might choose to increase leisure rather than surplus. Where there were difficulties in disposing of surplus or where the price that could be obtained was inadequate that strategy may have been optimal.

between 20 and 30 tonnes.⁸³ The construction of such units requires accessible markets which can consume this surplus - these units have no purpose other than to produce exchangeable surplus - and the exchange of that surplus for money or goods which will be consumed at some place other than where it was produced. The production strategies are, at least in part, influenced by consideration of what can be disposed of in the accessible markets which requires knowledge of how those markets might behave and very likely knowledge of how other potential suppliers might behave, all of which implies an interconnected landscape. Landscapes constructed from these units are the least densely populated since they are designed for maximum labour efficiency. This strategy requires the controller of production to severely limit the number of workers consuming the produce of the agricultural unit. This strategy is difficult to apply under the manorial system where there is limited labour flexibility.

Social and political factors shaped the landscape and, together with the production strategies employed within the economic units, determined the amount of grain that was available for consumption outside those units. Where landowners largely fell into the lowest social grouping little or no surplus was available. Where landowners fell into the highest social grouping, about 400 kg to 600 kg of grain per hectare of farmland was available for export to nearby urban centres or beyond. Landscapes intermediary between these two extremes may be thought to have produced between 200 kg and 300 kg per hectare for external consumption.

2.5 Conclusions

My analysis indicates that agricultural practices and technology had developed to the extent that productivity lay in the range of 1,200 to 1,800 kg per hectare of land set to grain, where land management practices and natural conditions were reasonable, with significant variations outside that range possible because of weather. This range of numbers, which is critical to the evaluation of an economy for which the production of food was the main activity, is higher than the most recent sets of figures produced by scholars and in some cases substantially so.

⁸³ Using mean yields of 1,200 kg/ha to 1,800 kg/ha and assuming a workforce of 10 workers, including supervisors, consuming 300 kg of wheat each and allowing 20% for seed and wastage. The surplus being considered here is the excess of the grain produced on the farm over the grain consumed on the farm. It is not economic surplus. There are other charges including taxes and, for tenanted farms, rent.

This level of productivity is about twice that achieved in Neolithic times and about half that achieved in modern times. While much of the development seen in recent times lay beyond the reach of the imperial Roman economy there were opportunities for development which were not accessed. The writings of the agronomists suggest that an earlier scientific approach based on systematic trial and error and exploitation of discoveries made through that process had either not been widespread or had been abandoned.

More significantly, this level of productivity is substantially higher than was achieved in medieval times and it was not until about the middle of the eighteenth century in Europe that productivity again reached the levels that were achieved in the time of imperial Rome. This reduction in the efficiency of food production had consequences for the level of urbanisation, as will be discussed in the next chapter. I attribute the decline in productivity to a degradation in land and crop management practices which followed on from a change in the structure of land ownership. Most land in the time of imperial Rome was owned outright by private landowners and these private landholdings varied in size from the very small to the large.⁸⁴ Land was freely available as an investment asset class and an outlet for savings. This structure, together with considerable labour flexibility, incentivised and favoured those with superior land and crop management skills and who applied and developed those skills. Good farmers could prosper and increase their wealth. The flexible land ownership structure that applied allowed for a range of production strategies including those focused on pure production for external consumption. Many of these production strategies were viable only where there was access through trade networks into markets. The later decline of urbanisation and the fragmentation of trade networks restricted the range of production strategies. These aspects are covered in later chapters.

The landowning structure that followed was inflexible. Land was not a marketable asset and large landowners had duties to an overlord. Most of those who worked the land either had rights to the use of that land but no ownership or no rights to it at all. Labour mobility was severely constrained and some of the production strategies previously available could not be applied. There was limited incentive to maximise returns. It was not until the middle of the eighteenth century, when land ownership and trade networks had reverted to a structure similar to that which prevailed in imperial Roman times that productivity reached the levels previously seen.

⁸⁴ The large imperial patrimony was run as if it were a private enterprise.

CHAPTER 3: Possible and actual levels of urbanization and the implications for social interaction and productivity.

3.1 Introduction

In the last chapter I considered the process of agricultural production, limiting the analysis to one grain type in order to enable comparison be to easily drawn between the economy of imperial Rome and other economies. I discussed the levels of grain surplus that could be produced from landscapes populated under different landholding structures. As noted in the last chapter grain has the attribute that it, when stored properly, does not deteriorate over short periods of time and thereby enables a separation between the points of food production and the points of food consumption. The separation between the points of production and consumption enables large numbers of people to live in close proximity to each other which has effects on human interaction. I will present in this chapter methods for relating the agricultural surplus quantified in the previous chapter to the levels of urbanisation seen in imperial Rome and I will also directly relate these levels of urbanisation to the overall levels of economic output. I will show that, although the economy was largely agrarian when measured by the numbers of people involved in food production, the larger part of the economic output of the economy of imperial Rome derived from urban activity.

3.2 The debate

Rostovtzeff believed that the cities of the Roman Empire resembled those of modern times and that the city contained a bourgeoisie who acted as capitalists.⁸⁵ Weber defined a city by the characteristics that the residents depend on commerce and trade rather than on agriculture for their income and by its having a market centre. He classified cities into three types. In the *consumer city* the large consumers derive their income from rents or state revenues. In the *producer city*, which he considered the modern type, the income derives from the making of product which is sold elsewhere. In this city the large consumers include entrepreneurs and the mass consumers are labourers and craftsmen. The third type is the *merchant city* in which the large consumers are traders who make their money on the local markets.⁸⁶ Finley took the view that the contribution of manufacturing activities to the urban centres in the ancient world was negligible.⁸⁷ The most extreme example of this was Rome which he described as

⁸⁵ Rostovtzeff (1957) 93, 142.

⁸⁶ Weber (1978) 1212-8.

⁸⁷ Finley (1999) 139.

‘the complete parasite-city’, in the sense that the imports into the city were paid for by income derived from land, trade or booty rather than from production.⁸⁸

Whittaker, writing in 1990, also believed that Weber’s assumption that the majority of the income of the urban consumers came from rural rents rather than commerce was correct.⁸⁹ Mattingley in 1997 claimed the consensus view of urbanisation to be that, other than in the case of cities such as Rome and Alexandria, urban centres depended on their immediate hinterland through rents and taxes, and that regions were self-sufficient. This view implies that manufacture and inter-regional trade were limited.⁹⁰ Parkins writing in 1997 followed the view that the urban economy relied on agriculture.⁹¹ Erdkamp writing in 2001 developed the consumer city model to be a consideration of the reciprocity or non-reciprocity of the relationship between the food producing sector and all other sectors of the economy. He believed that the economy is to be read as one in which the non-reciprocal relationship between food and non-food production sectors dominates and characterises the economy. Influencing this belief is assumption that the productivity of labour and soil was low.⁹²

3.3 Urban structures

The following table summarises the urban structure of the Empire.⁹³ The urban population structure was bi-modal, split between mega cities and small towns. These mega-cities have inhabitants who number between 500,000 and 1,000,000 and the rest of the population is almost entirely located in small towns with inhabitants numbering fewer than 25,000. There is hardly anything in between. The mega-cities are Rome, Alexandria, Carthage and Antioch and between them account for over 50% of the total urban population. This structure reflects the features of the transport system as will be discussed.

I will use two measures of urbanisation. The first is simply the number of people who lived in towns or cities and the second is that number expressed as a percentage of the total number in the population. While both of these quantities present estimation difficulties the first is easier

⁸⁸ Finley (1999) 130.

⁸⁹ Whittaker (1990) 112.

⁹⁰ Mattingley (1997) 210.

⁹¹ Parkins (1997) 87.

⁹² Erdkamp (2011) 344, 355, 352.

⁹³ The estimates are taken from Wilson (2011). The approach used to estimate the size of the urban population is based on the area contained within the walls of the urban centre, with detectable internal uninhabited land excluded and with detectable external inhabited land included, multiplied by a population density factor. Sense checks were also applied.

to quantify since it does not require any estimates of the total population.⁹⁴ The estimation of that quantity for imperial Rome is difficult because much of the population lived in rural settlements and in buildings which have left no trace and although there were censuses these are problematic because of difficulty in interpretation of results. On the other hand, comparison of the first measure, total urban population, in two economies allows comparisons to be made of the extent to which those economies supported activities which require an urban context.

⁹⁴ The uncertainties in the estimation of city and town sizes and in the estimation of the numbers of towns is not overwhelming. Morley (2011) 144-50 gives an estimation of low and high figures for urban populations. The central estimate in Morley (2011) is taken from Morley (1996) 182 and is 1.5 million which corresponds to the Wilson (2011) figure. The range around that figure given in Morley (2011) is from -25% to +25%. This level of uncertainty is such that the conclusions in this chapter are unaffected by reasonable variations in assumptions as to levels of urbanisation in Roman Italy.

By way of comparison the following table shows how the urban structure of England changed over the period between the fourteenth and nineteenth centuries.⁹⁵ In the period up and until the end of the seventeenth century the level of urbanisation, as measured by the number of people living in urban centres, was materially below that of imperial Rome. Judged simply by the numbers of people who lived there and compared with urban population estimates for imperial Rome, London through this period was an unremarkable town. The second city of England, York, was comparable in size to Pompeii.⁹⁶ Scarcely any Roman province had the low level of urbanisation that applied in England in the period before the eighteenth century. The total urban population of England in the fourteenth century was probably similar to that of Crete in imperial Roman times and in the sixteenth century was probably comparable to that of Sicily. By the middle of the eighteenth century the urban population of England started to approximate to that of imperial Roman Italy but not to that of imperial Roman Egypt. By the middle of the nineteenth century the structure had radically changed.

⁹⁵ The figures shown for York in the sixteenth, eighteenth and nineteenth centuries are taken from Tillott (1961) pages 121, 212 and 283 respectively. The population of Bristol for the seventeenth century is for 1600 and is taken from <http://www.localhistories.org/population.html> retrieved 24 August 2015. The figures for the fourteenth, sixteenth and seventeenth centuries for Sheffield are for 1379, 1600 and 1692 respectively and are taken from <https://www.sheffield.gov.uk/dms/scc/management/corporate-communications/documents/leisure-culture/libraries-copyright/archives/research-guides/population/Population-Statistics-Study-Guide--Word-363-KB-.doc> retrieved 25 August 2015. The figures for Manchester for the sixteenth and seventeenth centuries are taken from <http://www.localhistories.org/manchester.html> retrieved 25 October 2015. The sources of the remaining figures are as follows: The figures shown for the fourteenth century for Bristol, London and York are for 1377 and are based on Russell (1948) 273-8 which summarises the poll tax enumeration of 1377. The poll tax did not include all parts of the population and I have added a loading of 55% to reflect the exclusion of children, indigent and untaxed persons following Russell (1948). The figures for the sixteenth century are for various dates in the sixteenth century and are taken from Patten (1978) 103. The figures for the seventeenth century are for various dates in the seventeenth century and are taken from Patten (1978) 106. The eighteenth and nineteenth century figures are for 1750 and 1850 respectively and are taken from Mitchell (2007) 75-7. There appears to be very limited population data for the fifteenth century. The decline in London population in the seventeenth century was caused by plague. The 'Great Plague' of 1665 caused around 55,000 deaths. Hinde (2003) 107.

⁹⁶ Jongman (1988) 108-112 surveys the population estimates of Pompeii and concludes that an urban population figure of 8,000 to 12,000 for the town is 'plausible but high'. The range of population estimates is -25% to +25% of the central estimate. The level of uncertainty identified here is identical to that indicated by Morley (2011) for Italy as a whole. He assumes that the economic catchment area of Pompeii is about 20,000 hectares. Note that this catchment area corresponds to a radius of something under 10 kilometres and at this size a population of 10,000 is just about supportable from its contiguous hinterland, given the grain yields that were achievable. See later in this section. Additional population support could be provided through port structures.

Table 3.2 Illustration of structural changes in the English urban landscape

1: City	2: 14th century	3: 16th century	4: 17th century	5: 18th century	6: 19th century
London	36,000	90,000	34,000	675,000	2,685,000
York	11,000	8,000	12,000	17,000	36,000
Bristol	10,000	12,000	20,000	45,000	137,000
Birmingham		2,000	3,000	24,000	233,000
Liverpool		1,000	2,000	22,000	376,000
Manchester		4,000	5,000	18,000	303,000
Leeds		3,000	6,000	16,000	72,000
Sheffield	2,000	3,000	4,000	12,000	135,000

The overwhelming majority of towns enumerated in the 1377 poll tax had populations which numbered in the few hundreds. Indeed a large town had a population of about 800.⁹⁷

The process of urban growth is influenced by inwards migration.⁹⁸ The growth rates seen for most of the towns and cities in Table 3.2, some of which increased in size tenfold or more,

⁹⁷ Russell (1948) 307. Wilson (2011) 179 notes estimates for average populations for late medieval German towns of 500.

⁹⁸ So see, for example, Patten (1978) 121-135 for discussion of urban migration in the sixteenth and seventeenth centuries. Hinde (2004) 9 Table 1 shows that in the period 1841 to 1911 almost all of the natural increase in the

could not have been achieved without inwards migration. ‘The movement from country to town ... is motivated by the possibility, held out by the town, of considerable or even dazzling advancement; an opportunity held out to all, though in fact it is only a few of those who move who will achieve it. The call of Bow Bells to Dick Whittington – “Lord Mayor of London”! In their sound we have a symbol of enchantment.”⁹⁹ I believe that more humble aspirations were actual and sufficient motivation and were an indication of the relative value to the migrant of the urban and the rural opportunity.¹⁰⁰

A city such as medieval London with a population of less than 50,000 and a city such as imperial Rome with a population of about 1,000,000 are two very different places and represent two very different economic achievements. The economic structure in which the medieval city of London was embedded did not produce infrastructure comparable to that of imperial Rome. That infrastructure included water and food supply physical and institutional infrastructure, bath, sewage systems, ports, roads and markets.¹⁰¹ The very fact of the size of the city of imperial Rome is in any case an important economic variable. It is indicative of the existence of economic infrastructure but it is also indicative of economic activity, if for no other reason than that the people in the city had to support themselves.¹⁰² The grain dole in Rome was insufficient guarantee – it met a man’s grains needs, but not the rest of his diet, it could not support his dependents, and housing was not provided.

3.4 The relationship between agricultural surplus and urbanisation

To illustrate the relationship between agricultural surplus and urbanisation I use a variable **S**, which I define as the grain surplus produced in rural areas divided by the number of rural inhabitants and expressed as a percentage of a man’s annual grain consumption. **S** is the excess of the grain generated in food producing areas over the grain consumed in those areas whether consumed by agricultural workers or their dependents.¹⁰³ As described in Chapter 1 the production units on which this rural population lived ranged from small holdings which

rural areas – because of the excess of births over deaths – was lost to migration into the urban centres and the colliery districts. One third of the increase in the population of London in the period was from inward migration. The table is based on Lawton (1968), Cairncross (1949) and Cairncross (1953).

⁹⁹ Hicks (1969) 134-35 quoted by Patten (1978) 126. There was the additional benefit of freedom in medieval England. A villein who escaped to a borough for a year became free.

¹⁰⁰ Laurence, Cleary and Sears (2011) 7 notes the attractiveness of opportunities in the Roman city relative to those in the countryside.

¹⁰¹ In early nineteenth century London sewage was disposed of through the night trains.

¹⁰² Weber (1978) 1218 believed that in Antiquity, unlike in the Middle Ages, the urban citizen supported himself from his own land. This cannot be true of a city the size of Rome.

¹⁰³ I have assumed that grain consumption by livestock was minimal and that pasture and hay were used.

barely covered the food needs of those who occupied them to large holdings entirely focused on the production of food for sale. In all cases I consider that mixed strategies were followed on the production units which were used to grow a mixture of arable crops and in some cases, particularly for the larger units, also included the rearing of livestock. The surplus S is a function of the following variables

- Agricultural yield
- Rural population density/landholding structures
- Accessibility of markets/transport system
- Size of markets
- Random fluctuations

The higher the agricultural yield the more surplus is produced. If the rural population is reduced to the extent that those who work the land do so more efficiently or have fewer dependents then agricultural surplus increases because the amounts of grain consumed on the land rather than in the city is reduced. Surplus is produced only if the controllers of production believe that it can be disposed of profitably. Where access to markets is increased then controllers of production will have an incentive to select production strategies which increase surplus. Equally where there are hindrances to the transport of surplus from the production units to markets then some available production strategies which lead to higher surplus will not be selected since their implementation would generate surplus which cannot be sold. Increases in the size of accessible markets will tend to change selected production strategies such that surplus will increase. Finally surplus S varies randomly around a mean as weather conditions change year on year.¹⁰⁴

As demonstrated in Appendix 1 there is a clear link between the maximum urbanisation level and the level of grain surplus S .¹⁰⁵

¹⁰⁴ This random variation means that it is close to impossible to follow an agricultural production strategy which does not produce a surplus, on average. This is because large variations in the amount produced in the economic unit are expected, because of weather variations. If only enough is produced in an average year to be sufficient for the needs of those who live on the economic unit, then in many years they will be expected to starve and the economic unit collapses. The amount of production needs to be such that in most cases there is a surplus and in adverse conditions there remains sufficient produced or saved for the unit of production to continue. We expect, therefore, that the amount produced exceeds, in most years, the consumption needs of the producers. For simplicity in what follows I will consider approximate the surplus output from subsistence units as averaging at 0, although in practice it is a small positive number.

¹⁰⁵ I approximate the urbanisation level as the percentage of the population not involved in food production or not directly dependent on those who are. Urbanisation is maximised relative to a given surplus when all that surplus is consumed in the urban centre.

$$\text{Maximum urbanisation percentage} = S/(1+S)$$

Formula 3.1

S is a measure of potential urbanisation.¹⁰⁶ As surplus increases (i.e. as any of the above factors moves in a favourable direction) then higher levels of urbanisation are possible. The essential dynamic is that in the long run rural dwellers only produce food surplus to rural requirements if there is someone who will consume it. They do not work to produce food with the intent of letting it rot. This dynamic forces the relationship seen in Formula 3.1.

The following table illustrates the relationship between the structure of the agricultural landscape and the level of urbanisation.¹⁰⁷ I have considered variations based around small farms, producing very limited agricultural surplus, and large slave estates. The subsistence and slave production units are assumed to follow diversified production strategies. They do not just produce grain. The subsistence units are also assumed to engage in weaving and other domestic activities, particularly during agricultural downtime but to have limited degrees of specialisation and therefore limited exchange between subsistence production units. The slave estates maximise surplus because there are no dependents to consume surplus and because the labour force can be minimised to the level sufficient to work the estate given the labour technology available. Alternative variations using other units, such as moderately sized tenanted estates, are possible. The broad conclusions remain the same. The level of urbanisation is the average over the landmass across which the grain circulates. Column (1) shows a description of a rural landscape. Column (2) then shows what proportion of the agricultural landscape might have been covered by slave estates. Column (3) then shows the

¹⁰⁶ In fact **S** is the ratio of the number of urban dwellers to the number of rural dwellers since formula 3.1 implies $S = u/(1-u)$ where **u** is the urbanisation percentage.

¹⁰⁷ The parameters used are based on the work set out in Chapter 1. For the subsistence units I have used an agricultural yield at the low end of the range established in Chapter 1 because I expect such units to be inefficiently run. For the slave units I have selected an agricultural yield at the middle of the range established in that chapter. Appendix A illustrates these calculations. In Table 3.2 the required urbanisation for a rural landscape with a limited number of slave production units was given as 8%. The assumptions used are based on the analysis of Chapter 2. Here is an illustration of the assumptions and the calculations for one line in the table. The figures in the other lines are similarly derived. **Assumptions:** The subsistence units are assumed to be 5 hectares and have 6 people on each unit. The slave production units are assumed to be 50 hectares in size of which 25 hectares is set to grain and to have 10 people on each unit. The grain yield is assumed to be 1.5 tonnes per hectare and the consumption needs of a man are 0.3 tonnes per annum and 20% of the grain output covers seed and wastage. **Calculations:** If the agricultural landscape measures 500 hectares and it is 90% covered in subsistence units then there are 90 subsistence units and 1 slave estate. The total rural population is 550 (90 subsistence units * 6 people per subsistence unit + 1 slave estate * 10 people per slave estate). The surplus produced on the subsistence unit is 0. The surplus produced on the one slave estate is 27 tonnes (50 hectares x 50% set to grain * 1.5 tonnes per hectare * 80% not wasted or for seed – 10 workers * 0.3 tonnes grain consumed per worker) which is sufficient to support 90 urban dwellers (27 tonnes / 0.3 tonnes per man). **S** is 0.16 (27 tonnes / 0.3 tonnes per man year / 550 rural dwellers). Urbanisation is 0.16/(1.16) = 0.14. Alternatively the total population is 640 of which 90 are urban and the urbanisation percentage is then 14%.

corresponding maximum urbanisation level supported by that landscape. That maximum is reached when all agricultural surplus is consumed in the urban centres.

Table 3.3 Relationship between the rural landscape and urbanisation

1: Description of landscape	2: Percentage of worked landmass set to pure production estates	3: Maximum urbanisation supported by landscape
Exclusively small farms	0%	0%
Limited numbers of production estates	5%	7%
Small number of production estates	10%	14%
Moderate level of production estates	20%	26%
Level of urbanisation not achieved	40%	47%

Table 3.3 indicates that levels of urbanisation of between 20% to 30% of the total population could be achieved in the time of Imperial Rome with landscapes which were largely set to small agricultural units producing little beyond the basic needs of those who lived on them. This was because agricultural productivity was high and a relatively small part of the landmass devoted to pure production for trade could generate enough agricultural surplus to support large urban populations.

To achieve high levels of urbanisation there is a need to have high agricultural surplus and it is difficult to achieve high surplus without the level of labour efficiency produced by slave units or the equivalent. The choice to apply a highly efficient labour strategy depend on the existence of markets with accessible demand. The relationship between these two variables of urbanisation and surplus is mutual. Increased urbanisation may lead to changes in production strategies or changes in available surplus may lead to increased urbanisation. The change is

likely to be incremental and, I believe, demand led. It can also be reversed. If the town size falls then there is reduced demand and production strategies which were previously viable no longer work. Pure production units, for example, are less viable and estates focused on production for own consumption and more limited trade may take their place. Agricultural surplus need not be marketed into urban centres; rural production units trade between each other. However, the level of demand supported in such a trading structure is substantially below that supported by trade into urban centres with population densities much higher than in the countryside. In any case, trade between rural production units will be limited if they are producing the same commodities.

The essential dynamic here is one of changing population structures and changing demand. The process of moving from a landscape which is wholly set to subsistence units to one which is more dominated by pure production units is a process of rural depopulation as well as a process of increased surplus production. As people leave the country where the population is either almost entirely engaged in food production or is dependent on those so engaged and move to the urban centres they need to engage in new economic activities. The extremely limited state support for the poor meant that these entrants had to find work or starve.¹⁰⁸ Their demand for work pushes towards increased economic activity and individuals migrating to the cities increase capacity to produce goods and services. When this additional capacity is utilised, as new entrants to the city find outlets for their work, there is increased demand, as the new entrants have money to spend. The process can only take place where agricultural surplus is sufficiently high to support the increase in the urbanised population. This process can be seen in the rural depopulation around Rome that occurred in the last years of the Republic as the elite acquired and consolidated landholdings using the proceeds of war, displacing the existing rural population and replacing that labour with the slave labour also acquired through war. In the process those who had previously produced food became the consumers of food produced by others.¹⁰⁹ A similar process can be seen in eighteenth century England with the enclosure of lands. As noted earlier, medieval land-holding structures were not capable of producing the high levels of surplus supplied by the imperial

¹⁰⁸ The grain dole was distributed in few cities and only to a part of the population and was never more than a man's need. It did not provide support for dependents.

¹⁰⁹ See Rosenstein (2008) 1-2, for example, for a short description of this view and for criticisms of that view including those which relate to the extent of plantation agriculture and the importance of small farms to slave farms. (Slave farms require casual labour which is supplied by small farms.) Nevertheless, some consolidation of agricultural production is necessary to generate the high levels of surplus needed for city formation. See Kehoe (2007) 553-7 for a discussion of estate production.

Roman structures and could not support the levels of urbanisation that had been possible earlier.

Urbanisation levels of 50% were possible given the grain yields that were possible but these levels could not be achieved, because of other limitations on the size of towns. The nature of the land based transport system, for example, imposed a limitation on the level of urbanisation, as is discussed in the next section of this chapter, and in effect contributed to the shape of the rural landscape. At its simplest, large parts of the land mass were not sufficiently connected to urban centres to support production strategies focussed purely on external consumption.

3.5 The relationship between the transport system and the size of towns

The maximum daily range together with the amount of surplus available from agricultural units determined the size of towns that could be supported from their hinterlands. Hinterland is defined as the landmass contiguous to the urban centre from which the urban centre derives supplies. Part of this hinterland was needed for forest to provide fuel and building materials, part will have been inaccessible, infertile or simply uncultivated. I have assumed that 30% of the hinterland can be set to farmland by way of illustration. The following table explores the implications of agricultural production and the range of the transport system for urbanisation.¹¹⁰ Column (1) of Table 3.4 shows a range of average surplus amounts that might be produced in the hinterland of a town, given agricultural yields that were achievable in imperial Roman times. Note that the figures shown in column 1 as grain surplus are amounts of grain produced less amounts consumed on the production unit. Column (2) then shows how many urban dwellers could be supported by that hinterland if it extended out 5 kilometres from the town. Column (3) then shows the larger town that could be supported by a larger hinterland. So a town with a hinterland which has a radius of 5 kilometres producing an average surplus of 100 kg per hectare of agricultural land has a population of 1,000. Column (4) provides a description of the landscape needed to produce that level of agricultural surplus.¹¹¹ The analysis supporting the range of surplus values is provided in Chapter 2. I have used 10 kilometres as a maximum hinterland radius because of the daily range of the land based transport system. See Appendix A2. I believe that most of the goods

¹¹⁰ Population is calculated as $\text{Surplus} \times (\pi \times \text{hinterland radius squared}) \times 100 \times 30\% / 300$, using 100 to convert from km^2 to hectares and assuming that 30% of the hinterland was available for agricultural land. The average annual wheat requirement per capita is taken as 300kg. Results are rounded to the nearest 500.

¹¹¹ Farms of about 25 hectares produced a surplus of between 0 and 5 tonnes in an average year. Slave estates of about 50 hectares produced a surplus of between 20 and 30 tonnes in an average year. See Chapter 2.

transported into the towns and cities by donkey or animal drawn wheeled transport were carried in on day round trips and the men and animals transporting the goods in did not overnight there. A donkey can travel up to about 20 kilometres in a day and a day's round trip is therefore a maximum of 10 kilometres and probably less than that if we allow for unloading and waiting times. All goods transported in overland from beyond 10 kilometres had to be taken in by men and animals who overnighted in the town and had to be accommodated there.¹¹²

Rural production units which were within 5 or 10 kilometres of a town or of a port or had river access to a town or a port could pursue production strategies which were directed towards meeting urban demand and could be pure production units.

¹¹² Note Finley (1999) indicates an average distance from the local market to the edge of the area which contributed to the market of between about 6 and 8 kilometres.

Table 3.4 Relationship between land locked town size and hinterland radius

1: Average grain surplus kg/ha	2: Town size Hinterland radius 5 km	3: Town size Hinterland radius 10 km	4: Description of landscape
100	1,000	3,000	Significant number of small farms
200	1,500	6,500	Moderate number of slave estates
300	2,500	9,500	Significant number of slave estates
400	3,000	12,500	Very largely slave estates

The level of agricultural productivity in the time of imperial Rome towns was such that towns of up to 10,000 could be supplied from their land based hinterland. Given yields and land transportation constraints town sizes above 10,000 **require** at least some of the following: significant sea or river port infrastructure, landscapes which have few small holdings, landscapes which do not have the need for fallow, are highly fertile or which have very few inaccessible areas. Even in the middle of the eighteenth century such towns in England were rare. The construction of railways substantially altered these relationships because it changed the size of the hinterland. Increasing the radius of the hinterland from 10 kilometres to 50 kilometres, for example, changes the maximum town size from about 10,000 to about 250,000. Production units which were more than 10 kilometres from a town or port become accessible. More widespread consolidation of landholdings becomes economic as does the implementation of improved land management.

I have highlighted the row which shows 200 kg per hectare for which the corresponding town sizes are 1,500 and 6,500. This is the absolute maximum that could have been produced under medieval yields. This level of surplus requires that the agricultural landscape is populated by labourers only, that there are no dependents and that those labourers are worked to the full. In practice the maximum town size that would be supported from its hinterland in medieval times was substantially below this because of labour structures. Additionally the

small sizes of towns that could be accessed by travelling overland from the place of production to the place of consumption meant that there was limited incentive to improve land productivity. The local demand did not exist. The size of the hinterland was small and the agricultural surplus produced was negligible. The yield levels achieved in eighteenth century England and in imperial Rome are comparable, as discussed in Chapter 2, and the transport systems are similar. The maximum town sizes are therefore also similar.

The following table illustrates the supply characteristics of the towns.¹¹³ At the higher levels congestion of the overland transport limits growth and water based transport begins to require infrastructure which needs municipal or state involvement in direction and finance. This infrastructure also needs continuing maintenance and starts to generate a relationship between town size and state or municipal finances and stability.

Column (3) shows the number of ship journeys per month required to supply the various towns with their full requirements of grain. Except at the highest town size shown river or sea ports have very low levels of activity and investment in port infrastructure was unlikely to have been economic for towns of these sizes, except to the extent that infrastructure serves multiple towns. Towns were therefore either small and supported by their hinterland or very large and supplied through substantial port infrastructure. This bi-modal distribution of town sizes is evidenced in Table 3.1.

¹¹³ Land based figures calculated on the assumption that per capita consumption of wheat is 300 kg and that a two animal ox cart will carry 1 tonne of grain. The sea based figures are based on the assumption that a ship delivers 100 tonnes of wheat. I have used 100 tonnes as illustrative out of a normal range of ship sizes of 60 tonnes to 400 tonnes. I have taken a figure at the lower end of the range because I am considering relatively small towns. These figures for ship sizes are based on Rathbone (2003) 199- 201 which is cited in Scheidel (2011). The transport parameters are taken from Appendix 2.

Table 3.5 Comparison of land and sea supply

1: Town size	2: Land supply requires this number of daily ox cart arrivals	3: Sea supply requires this number of monthly ship arrivals at port
2,000	2	
5,000	4	2
10,000	8	3
25,000	21	6
50,000	41	12

3.6 Relationship between urbanisation and total output

Scheidel and Friesen in 2009 summarise quantifications of the size of the Roman economy, measured by estimates of GDP.¹¹⁴ They compare different views on what the price of wheat was and what the population of the empire was. The structure of that estimation process rests on the assumption that the average wheat consumption was variously 175 kg, 200 kg or 250 kg per person per year.¹¹⁵ The GDP estimates per capita are just under 3.5 times this quantity.¹¹⁶

It is possible to construct an alternative model which relates total output to food production but makes allowance for views on urbanisation and the type of activity that took place in towns and cities, as I will do now. The first part in the process is to split the thinking into rural and urban. The process of surplus production in the city is different to that in the

¹¹⁴ Scheidel and Friesen (2009) present three approaches. The expenditure approach, the income approach and a comparator approach which assumes that the economy is comparable to Byzantium.

¹¹⁵ Hopkins (1980) adjusted uses 200 kg, Temin (2006a) uses 175 kg, Maddison (2007) and also Scheidel and Friesen (2009) uses 250 kg. My figure is 300 kg. The Hopkins adjusted figure of 200kg is his 250 kg – which is all food consumption – adjusted by his factor of 20% to get his implied figure of 200 kg for only wheat consumption.

¹¹⁶ The ratios are 3.34, 3.38, 3.50 and 3.40 for Hopkins (1980) adjusted, Goldsmith (1984), Temin (2006a) and Maddison (2007) respectively. Hopkins adjusted figure is his 334 kg – which is his absolute minimum –adjusted by his factor of 100% to get his implied figure of 668 kg for all consumption. See Jongman (2006) 237-8 for commentary on the value of estimates of GDP. Rathbone and Temin (2008) 374 suggest that per capita income was about 3 times subsistence. Hopkins (1980) gives a relationship between wheat consumption and total food consumption of about 80%. Using this relationship all these figures are broadly equivalent to 3 times subsistence.

country. The rural sector is essentially involved in food production and its total output is the total amount of food consumed. The urban sector is involved in the production of a wide range of goods and services and typically has higher value added. The total economic output of an economy can be expressed in the following formula, the derivation of which is set out in the Appendix 3.

Total output = rural output x (1+ urbanisation proportion x relative urban productivity) Formula 3.2

This relationship has a number of consequences. In an economy where the urban centres are populated exclusively by pure consumers then urban productivity relative to rural productivity is zero and the total output of the economy is the food produced, or more generally it is what is produced outside the urban centres. If we form views on relative urban productivity then we can form views on overall economic activity.

The Figure 3.1 illustrates the modern relationship between per capita income and urbanisation for 180 countries for 2000. The correlation between the two variables is 0.80.¹¹⁷

¹¹⁷ Bloom, Canning and Fink (2008).

Figure 3.1: Modern comparison of income and urbanisation levels

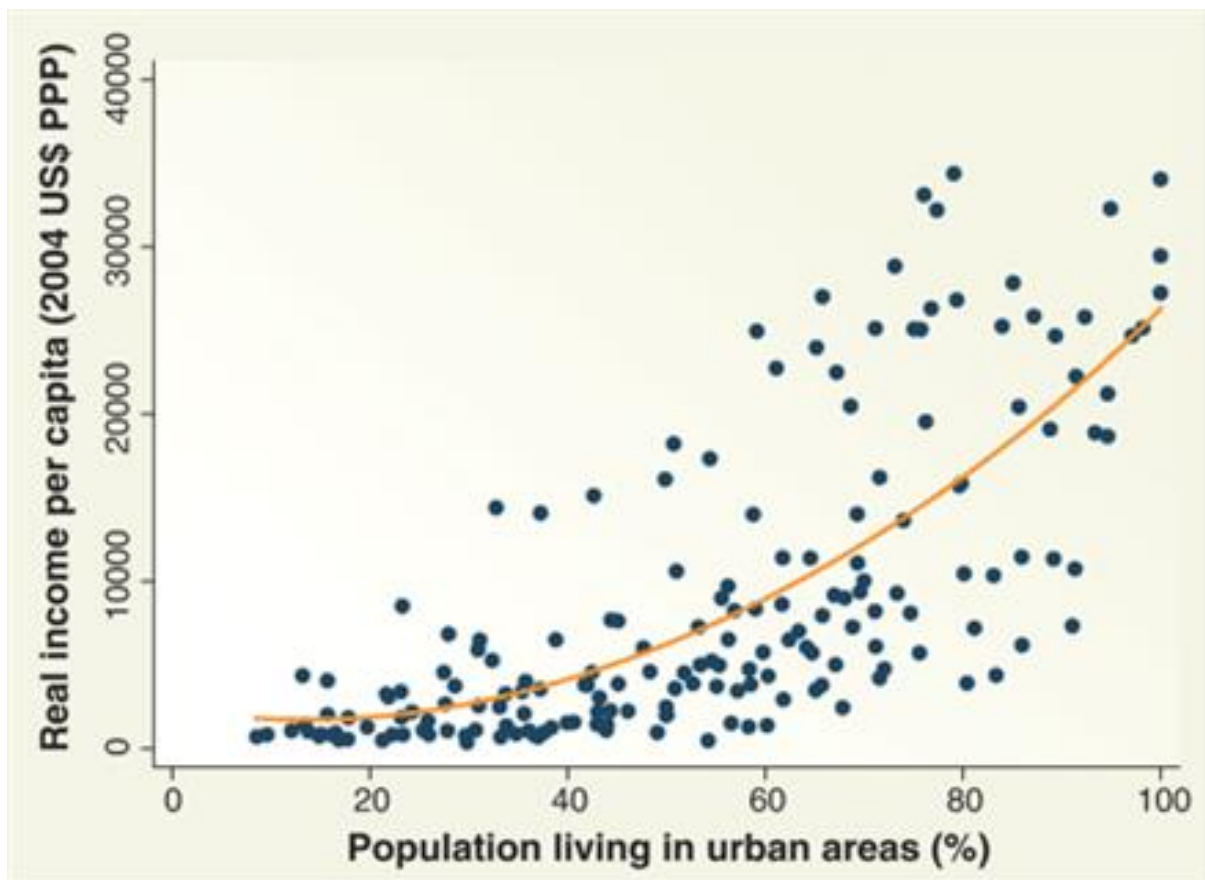
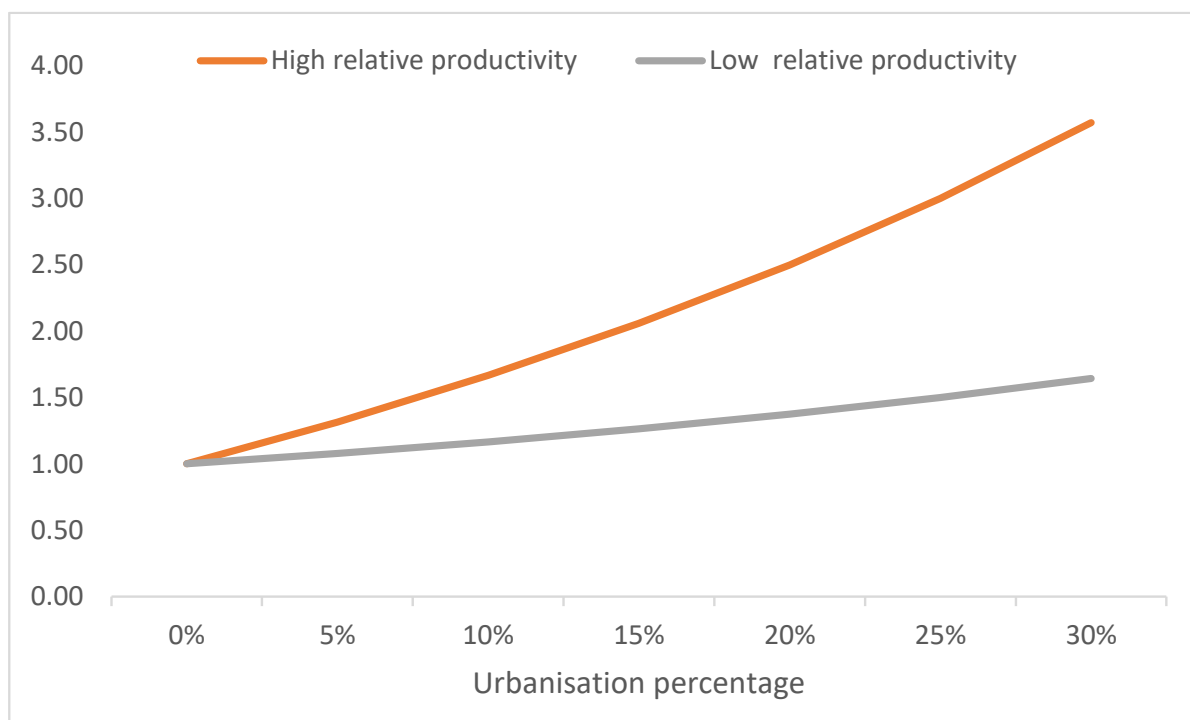


Figure 3.2 below shows the relationship between total output and urbanisation based on my formula 3.2. Total output is measured as multiple of food production. I show two lines. The higher line is where urban productivity is very high relative to rural productivity and the lower line is where urban productivity is moderately higher than rural productivity. In this context by higher productivity I mean the capacity to produce higher value added and not simply the amount of time worked. Towns and cities offer opportunities for labour to be more productive than rural labour which, at least in the time of imperial Rome, was very largely manual. There is more opportunity to develop and deploy specialised skills. I believe that the divergences shown in Figure 3.1 above reflect the difference in relative urban productivity in the different countries.

Figure 3.2 Total output as a multiple of rural output for different levels of urbanisation



There are essentially three ways in which total output can be increased. These are firstly through increases in the amount of food produced, such as by improvements in grain yields, but remember that improvements in the quantity of food surplus produced requires an increase in the numbers of people who consume but do not produce food.¹¹⁸ Changes in the types of food produced can lead to increases in total output where, for example, grain is replaced by higher valued items such as meat. Secondly, increases in urbanisation also lead to increases in output, provided that urban productivity is at least marginally higher than rural productivity. This is because for the same total number of workers the same quantity of food is produced and there is the additional urban output. Thirdly, total output is increased when urban productivity increases. This might be because of improvements in the skills of the urban population or reduction in urban unemployment, for example. Trade is also important because then the production of high value items with a limited local demand but extensive demand over a large area becomes possible and this increases urban productivity relative to rural productivity.

¹¹⁸ Other ways in which total output can be improved are through improvements in rural productivity (for example, the introduction of the combine harvester) with the displaced rural workers actively employed in the urban centres.

The estimates of GDP per capita already referred to suggest that total output is about 3 times rural output. At levels of urbanisation around 30% this suggests that urban productivity was about 4 times that of rural productivity. For much lower levels of urbanisation we cannot get to this overall level of income without assuming excessively high levels of relative productivity.¹¹⁹ In the next section I will consider the productivity of the elite and of the non-elite sections of the population.

3.7 Productivity

The following passage, from Cicero, is often cited as support for the view that the economy of imperial Rome was fundamentally different to that of modern economies.¹²⁰ The status considerations outlined here are seen as limiting economic activity.

Now in regard to trades and other means of livelihood, which ones are to be considered becoming to a gentleman and which ones are vulgar, we have been taught, in general, as follows. First, those means of livelihood are rejected as undesirable which incur people's ill will, as those of tax gatherers and usurers. Unbecoming to a gentleman too and vulgar are the means of livelihood of all hired workmen whom we pay for mere manual labour, not for artistic skill, for in that case the very wage they receive is a pledge of their slavery. Vulgar we must consider those also who buy from wholesale merchants to retail immediately; for they would get no profits without a great deal of downright lying, and verily there is no action that is meaner than misrepresentation. And all mechanics are engaged in vulgar trades, for no workshop can have anything liberal about it. Least respectable of all are those trades which cater for sensual pleasures: Fishmongers, butchers, cooks and poulterers and fishermen as Terence says. Add to these, if you please, the performers, dancers and the whole *corps de ballet*.

But the professions in which either a higher degree of intelligence is required or from which no small benefit to society is derived – medicine and architecture, for example, and teaching – these are proper for those who social position they become. Trade, if it is on a small scale, is to be considered vulgar; but if wholesale and on a large scale,

¹¹⁹ Temin (2006) 135 notes the urbanisation level of 30% for the early Roman Empire estimated in Hopkins (1978) 68-69 and notes that this is similar to that of the 1700 economies of the Netherlands, Italy and Spain. He also notes that it is routine to use the level of urbanisation as an indicator for per capita income. I note, however, from the data in Table 3.1 that levels of urbanisation in themselves are a relatively poor explainer of per capita income.

¹²⁰ Finley (1999) 41-4.

importing large quantities from all parts of the world and distributing to many without misrepresentation, it is not greatly to be disparaged. Nay, it even seems to deserve the highest respect, if those who are engaged in it satiated, or rather I should say, satisfied with the fortunes they have made, make their way from the port to a country estate, as they have often made it from the sea into port. But of all the occupations by which gain is secured, none is better than agriculture, none more profitable, none more delightful, none more becoming to a freeman.¹²¹

The passage of Cicero presents a view which is similar to that of the upper middle class characters in Jane Austen's *Pride and Prejudice*, published in 1813, who despise those whose income is not drawn from land or money put out at interest. The novel was written when England's dominant global position was based on its industrial production and overseas trade.

A short pause followed this speech, and Mrs. Hurst began again. 'I have an excessive regard for Jane Bennet, she is really a very sweet girl, and I wish with all my heart she were well settled. But with such a father and mother, and such low connections, I am afraid there is no chance of it.' 'I think I have heard you say, that their uncle is an attorney in Meryton.' 'Yes; and they have another, who lives somewhere near Cheapside.' 'That is capital,' added her sister, and they both laughed heartily. 'If they had uncles enough to fill all Cheapside,' cried Bingley, 'it would not make them one jot less agreeable.' 'But it must very materially lessen their chance of marrying men of any consideration in the world,' replied Darcy. To this speech Bingley made no answer; but his sisters gave it their hearty assent, and indulged their mirth for some time at the expense of their dear friend's vulgar relations.¹²²

'There were some very strong objections against the lady,' were Colonel Fitzwilliam's words, and these strong objections probably were, her having one uncle who was a country attorney, and another who was in business in London.¹²³

The Cicero passage is in essence about the vulgarity of certain occupations as seen by one part of the population.¹²⁴ It is not descriptive of the attitudes of the parts of the population

¹²¹ Cicero *De Officiis* 150-1. Loeb translation.

¹²² Austen *Pride and Prejudice* Chapter 8.

¹²³ Austen *Pride and Prejudice* Chapter 33.

¹²⁴ Joshel (1992) 82 suggests that this attitude may come from the fear that such work produces wealth, which underpins social standing. In other words that Cicero's attitude is an attempt to exclude the socially threatening. I believe that the depiction of Trimalchio can be similarly read. The satire of the immensely rich non-elite arriviste Trimalchio was written in a time when trade infrastructure, described in the next chapter, was being significantly expanded and elite fears may have been becoming more acute.

important to economic activity. The contribution that the non-food producing parts of the population make is what most drives economic development. The more painters or teachers there are relative to farm workers the more developed the economy. The incentives that non-elite individuals have to develop skills effects economic growth. In this respect what Cicero says or thinks is neither here nor there. His status group did not monopolise, certainly in the time of imperial Rome, the opportunities to create wealth.¹²⁵ Also, as will be discussed later, both elite and non-elite members saw participation in such economically progressive activities as the building of infrastructure which supported trade as enhancing their status.¹²⁶

The attitude described, because it favours land as an investment, implies that the surplus returns from land are reinvested back into land.¹²⁷ To the extent that available land was limited and not expanding through conquest, this tended to force consolidation through the buying up of smaller units.¹²⁸ This process in turn tended to force increased urbanisation, at least to the extent that the consolidation increased the level of surplus production, and the development of the crafts which Cicero despises, as an occupation for himself.

In both of these passages those who had land are decrying those in other occupations on whom they actually depend, whether as providers of goods and services they need, or as the purchasers of the output from their estates.¹²⁹ The characters in *Pride and Prejudice* are set in an economy which is more than just an agricultural food producing land based economy. Significant wealth was being created in trade, commerce and manufacturing. The status attitudes illustrated in the passages are indicative of the attitudes of a subsection of the population who live on the income from capital. Neither passage indicates that this was a general approach within these two economies and the evidence is that it was not prevalent in either.

¹²⁵ This aspect is discussed in the next chapter, which deals with trade.

¹²⁶ See Chapter 6.

¹²⁷ I believe that the economic elite, whether the fictional arriviste such as Trimalchio or more established such as Pliny, did favour land as the repository of their wealth because of status considerations. Note that some at least of the nineteenth century great aristocratic fortunes rested on urban demand. The wealth of the Earls of Fitzwilliam was based on coal extracted from their lands and the great fortune of the Duke of Bridgewater derived from canals.

¹²⁸ The fact that land was not always available as an investment is illustrated by Pliny, *Ep* 10.54. Pliny writes to Trajan that the local administration has collected the debts owed to it and now has surplus funds. It is difficult to find land for sale and there are no borrowers willing to take money at the normal 12%. He recommends reducing the rate charged on borrowed funds. This passage can also be read as indicating an ability and willingness by Trajan and by Pliny to engage in negotiation, which Weber considered inconsistent with pre-modern status group behaviour. Pliny, *Ep* 10.55 shows Trajan indicating that the rate will depend on the number of borrowers.

¹²⁹ As noted by Holleran (2012) 6 the low regard for shopkeepers and retail activity continued into the twentieth century.

The approach adopted by Finley flows from social concepts. The social framework is important because the stories that people tell themselves and which they collectively believe are important in determining what happens in an economy.¹³⁰ Finley draws on sources such as the above passage from Cicero to support his belief that the economically significant agents in the ancient economy were averse to trade. He contrasts this with modern economies and believes that this view was powerful enough that the ancient economy in general should be considered as different in type to modern economies. It is, however, not sufficient to identify and use a story in the literary sources that someone told about what should or should not be done because then, as now, different stories are told by different actors. The story Cicero tells is not necessarily the story that the elite group generally told themselves nor is it likely to be the story that others, the freedmen, for example, told themselves and the stories told in the time of Cicero, when the conquest model still largely applied, might not be the stories that were told and believed at later times when conquest was no longer the main route to status and wealth. I do not believe that elite aversion to trade mattered very much in the period of this investigation or that it was any hindrance to economic activity.

In practice the elite are economically active, not least because they own land.¹³¹ To the extent that his estates are managed directly for him by his dependents and produce grain, wine and other agricultural product, then a member of the elite is in the food business and needs customers. The production of so many tonnes of grain and so many hectolitres of oil and wine beyond what he and his dependents can consume is only worthwhile if that surplus can be transformed into goods or services which he desires or somehow monetised. To the extent that his estates are let out, then he is in the property business and he or his staff have to supervise the drawing up of leases, the collection of rents and the protection of their assets.¹³²

¹³⁰ So see Harari (2014) generally.

¹³¹ The elite wealth is generally seen as being property based. Champlin (1991), 101 notes the extremely limited number of ways of becoming rich in the Roman world. Jones (2006), 253 and Kehoe (1997), 15 see agricultural land as the main investment of the elite. Rosenstein (2008) argues that the elite also derived income from sources other than land.

¹³² There was considerable variety in the strategies that were employed by elite owners of property, which ranged from delegated management by tenants who were expected to provide all the working capital to active management by owners of complex networks of estates. See Kehoe (1988) risk averseness and for limited investment by landowners in working capital. But see also Rathbone (1991) for active estate management. There was no institutional preference for any particular strategy. Tenancy, with regular flows of income to the landowner, was also not the only relationship that existed between the owner of land and the user of it; owners could capitalise their interests by selling rights to use their property on a range of different terms. Most of Book 7 of the Digest is given over to law related to these rights. There are analogies with more modern capital extraction by freeholders selling leases in their properties.

In this case too the availability of markets for the output of tenanted estates mattered to the landowner since the higher the demand for agricultural product the greater the potential return to the landowner and the lower the risk of tenant default and asset deterioration. Elite property owners were aware of these considerations. Pliny, for example, presents himself as aware of the investment characteristics of property. He knew that property values can fluctuate over time because of scarcity, legislative changes can lead to property booms, poor harvests and adverse economic conditions will depress rental income and excessive rents increase the likelihood of void periods; risk can be reduced by following a diversification strategy.¹³³ Pliny was as economically active as the executive chairman of a property company.

Income is produced from these assets only if there is consumer demand for agricultural products. This consumer demand only existed to the extent that there were individuals who had a need for grain, oil and wine, for example, beyond their capacity to produce for their own consumption. That part of the population which was self-sufficient in these goods or operated at a subsistence level contributed limited or no demand. The population of the urban centres provided demand as did those living in the country who adopted specialised agricultural strategies.¹³⁴ In practice, therefore, land as a source of significant income required urbanisation. Additionally, being a substantial landowner with no distribution network into the urban centres was to be poor. The economic interests of the elite lay in the development of roads, harbours and markets and the legal apparatus that protected that infrastructure. It also lay in increased urbanisation and trade. Moreover, the markets in which the output of the landed estates was monetised could only function if the consumers of that output were engaged in economic activity that had monetary reward. This requires an economically active urban population whose interaction with the landowning elite is not through payment of rent to the elite but in the payment of money in return for agricultural output produced under the management of the landowner or by his tenants. These non-elite members had to be engaged in some economic activity. Any economic activity - whether the

¹³³ Pliny, *Ep.* 10.8.5 argues the advantages of the landowner taking a hands on approach; See Pliny, *Ep.* 3.19 for an example of how difficult economic conditions can reduce revenues and therefore capital values by 40%; See Pliny, *Ep.* 6.19 for a view on a property boom; See Pliny, *Ep.* 3.19 for a discussion of property diversification and consideration of what makes for good agricultural yields. See Pliny, *Ep.* 9.3.7 and Pliny, *Ep.* 8.2 for examples of good management and Pliny, *Ep.* 6.3.1 for a description of how poor land management reduces capital values.

¹³⁴ For example, someone who sets all his land to vines and has demand for grain and oil.

production of common footwear or luxury shoes - was sufficient so long as it generated money.

The demand structure of the city was not the simple structure of an elite property owning group which supported a group of direct dependents, family members and household slaves, for example, and on which the balance of the population was dependent for the full extent of their income, less whatever was made available by the state. Rather it consisted of that elite group, including direct dependents, plus an economically active group which provided goods and services to the elite and which exchanged goods and services between themselves and with others in different locations.¹³⁵

Table 3.6 illustrates the wide range of occupations that existed in Imperial Rome. These are found on memorial inscriptions.¹³⁶ The list of occupations is not much different to the list of occupations to be found in eighteenth century London. In addition to the categories shown in Table 3.6 there are hundreds of occupational titles requiring a high degree of

specialisation.¹³⁷ These include: comic actor, rent collector, pearl setter, maker of felt footwear, dealer in cloaks, dealer in oil from Baetica, money changer, maker of women's shoes, caretaker of mirrors, agent in charge of household expenditures and clothes folder.¹³⁸

There was a high degree of specialisation. At Rome, for example, there were craftsmen whose occupation was in the making of eyes for statues. The production of a silver vessel could be carried out by a succession of different individuals specialised in different aspects of the manufacture carrying out their business in workshops which were integrated with each other.¹³⁹ There was also a high degree of clustering as evidenced in toponyms in Rome.¹⁴⁰

These occupations required urbanisation and indicate a monetised economic structure in which exchange takes place across all levels in society rather than being dependent on an economic or social elite. The tailor and the shoemaker exchange with each other and with the

¹³⁵ Kron (2012) 255-7 takes a similar view. Jongman (1988) 187-203 considers the relationship between elite demand and total population and concludes that almost all economic active agents produced for elite consumption or were engaged in payment of rent to elite members.

¹³⁶ Joshel (1992) 69 notes that these epitaphs show men and women expressing pride in occupations which the literature decries. This collection of occupations may present a biased description of the total set of occupations although I believe that any such bias is unlikely to be material. Monuments tended to be erected by senators, veterans and freedmen rather than by the free poor or non-elite. Occupations in which the free poor or non-elite were exclusively engaged are unlikely to appear in these lists. So, for example, all of the occupations in Table 3.6 are skilled. There are no citations for those engaged in unskilled labour such as waste disposal.

¹³⁷ Zimmer (1982) 17-53 gives a categorisation of occupations drawn from memorial inscriptions an depictions and a history of their evolution.

¹³⁸ Joshel (1992) 174-82.

¹³⁹ Holleran (2012) 26-7.

¹⁴⁰ Holleran (2012) 51-60.

butcher. The urban setting provides advantages to these workers. A tailor situated in a city could do his business from a workshop.¹⁴¹ As with many of these occupations the tailor needs raw materials and is best situated close to markets which supply those raw materials. The alternative models to be applied in the country are either an itinerant worker visiting rural dwellings or part time work carried out by inhabitants of agricultural production units. Neither of these alternative models offers the tailor the opportunities and benefits of working from an urban workshop and both of them impose additional costs.¹⁴² A craftsman situated in the country could only service those who lived within a day's return journey from where he lived. One situated in a city could have a potential customer base in the many thousands. Equally the concentration of tailors in the urban centres offers the buyer choice and more secure supply generating a preference on the part of buyers for urban centres. This buyer preference in turn motivates skilled workers to migrate into urban centres where the proximity of buyers will give them better opportunities than are available in hamlets. Many of the occupations shown in Table 3.1 are found in towns, not on farms, and it is impossible to carry out most of these activities to any worthwhile extent except in cities or towns. The table illustrates the opportunities available to migrants into the city.

¹⁴¹ There were typically leased by craftsmen. Holleran (2012) 46-8.

¹⁴² The itinerant worker spends time travelling rather than working, for example.

TABLE 3.6
LIST OF OCCUPATIONS

I) CONSTRUCTION	IV) BANKING
a) Carpenters and builders	V) EDUCATED SERVICE
b) Stoneworkers	a) Architects and surveyors
c) Stucco workers, mosaic workers, painters	b) Doctors and midwives
d) Contractors	c) Teachers
II) MANUFACTURE	VI) SKILLED SERVICE
a) Jewellers	a) Barbers, hairdressers, masseurs/euses
b) Metalsmiths	b) Doctors and midwives
c) Workers in ivory and glass	c) Teachers
d) Producers of cloth (including spinners, wool weighers, weavers, dyers, fullers)	VII) DOMESTIC SERVICE
e) Makers of clothing	a) Child attendants and nurses
f) Shoemakers	b) Bodyguards
g) Artisans in leather	c) Personal, room and table servants
h) Butchers and bakers	d) Cooks
i) Artisans of other products (including artisans of tools, pipes, furniture)	e) Provisioners
III) SALES	f) Caretakers
a) Dealers in food products (including dealers in meats, fish, produce, oil, wine)	g) Gardeners
b) Dealers in unguents and perfumes	h) Social organisers
c) Dealers in leather, cloth and clothing	VIII) TRANSPORTATION
d) Dealers in books and paper, etc.	a) Animal tenders and baggage handlers
e) Dealers in metal and marble	b) Runners and bearers
f) Dealers in slaves	c) Drivers and boatmen
g) Dealers, product unspecified	IX) ADMINISTRATIVE SERVICE
	a) Financial agents
	b) Administrators
	c) Secretaries, clerks, copyists

The following table summarises annual earning for various job categories as taken from Diocletian's price edict.¹⁴³ For each of these jobs the edict indicates that maintenance is in addition. By way of reference point the cost of a man's annual grain needs is 4,500 denarii.¹⁴⁴ The earnings are expressed as multiples of what was to be paid to farm labourers. Although I do not believe that the price edict had its intended effect of curbing the high inflation which had emerged in the third century AD I do believe that the relationships between the prices set out in it are useful. The relationships also seem reasonable in themselves. The more skilled workers are paid more than the less skilled.

Table 3.7 Relative earnings of various occupations

1: Job description	2: Relative earnings (farm labourer as base)
Farm labourer	1.0
Sewer cleaner	1.0
Muleteer	1.0
Wall mosaic painter	1.2
Shipwright of seagoing vessel	1.2
Carpenter	2.0
Baker	2.0
Picture painter	3.0

Farm labourers, muleteers and sewer cleaners, who needed little training, could hardly support themselves and a small family.¹⁴⁵ More skilled workers, such as carpenters and bakers, could just about support themselves and a small family. Men at higher levels of training, such as picture painters, could live well. With these levels of earnings differentials there were clear financial incentives to acquire marketable skills.¹⁴⁶ Equally there were incentives to migrate to the cities where those skills could most profitably be deployed. The relationship between earnings and annual grain needs indicates that for the higher skilled

¹⁴³ I have converted daily wages into annual earning assuming 250 working days in a year.

¹⁴⁴ 100 denarii for 1 army modius (= 6.7 kg) per the edict and an annual need of 300 kg.

¹⁴⁵ The daily rate for a farm labourer is 25 denarii and the price of wheat is 100 denarii communes for one modius or 6 kg. The cost for a farm labourer in Matthew 20:1-16 is 1 denarius per day.

¹⁴⁶ Slaves may have had an advantage in this process. Holleran (2012) 29 notes the training advantages that slaves as opposed to free citizens had in getting a training.

workers their earning capacity substantially exceeded their food needs. They could consume more than was produced by the agricultural estates of the rich. That excess demand was satisfied by urban production. I believe it is reasonable to assume that urban productivity was 4 times that of rural productivity in imperial Rome, and it may have been higher. This level of productivity gives per capita income which is broadly equivalent to the figures discussed in subsection 3.4.¹⁴⁷ Since overall output depends on the level of urbanisation and relative urban productivity then we can draw broad comparisons of the per capita total output over time. Scholarly consensus puts per capita income for the time of imperial Rome at about 3 times subsistence. In medieval England urbanisation was substantially below the level seen in imperial Rome and per capita income was accordingly much lower. I believe that in eighteenth century England per capita income had risen to the level seen in the time of imperial Rome.

3.8 Conclusion

As described in Chapter 2 the capacity of the economy to produce surplus grain for consumption in local urban centres or for export depended as much on ownership structures and on the production strategies that owners choose to follow as on agricultural practices and technology. Not all farmland was used to grow grain, and some of the grain was consumed by those who produced it. Landscapes populated by smallholdings produced limited surplus. Landscapes populated by large holdings produced surplus of between 400 kg and 600 kg per hectare of farmland, provided there were accessible markets with which owners needed to or wished to trade. Productivity is as much a product of the political and social environment as it is of technological capabilities.

This level of agricultural productivity together with the range of the transport systems was a critical limiting factor on urbanisation. Given the numbers quoted above for grain productivity, the range of the land based transport systems and reasonable assumptions as to producer strategies it follows that urban settlements of more than 10,000 could generally not be supported from their contiguous hinterlands. Settlements of that size and higher could only be sustained by river and sea based transport systems, which generally requires the development of infrastructure.

The relative importance of the various economic parameters can be illustrated by the following scenarios. Consolidating the hinterland to remove small holdings increases the ceiling from a

¹⁴⁷ Those multiples to subsistence are difficult to achieve without relatively high levels of urbanisation (more than 25%, say) and relatively high levels of urban productivity (more than 4 times rural).

population of 10,000 to a population of 20,000.¹⁴⁸ Doubling the transport range further increases this to 80,000.¹⁴⁹ A 50% improvement in crop yields increases this to over 100,000.

As also noted in Chapter 2, annual crop yields vary simply because of local changes in the weather. Local variations in yields of up to 30% are not rare for the empire as a whole. When reductions in yields of this level occur many smallholdings are unable to support the families living on them. However, the fall in output on the larger estates reduces surplus available for sale while leaving sufficient for those dependent on the estate to live. The towns supplied only by these estates in their contiguous hinterlands will, unless they have maintained stores, have insufficient to feed their inhabitants. The normal features of a famine prevail. Grain is produced but the poor starve. Towns supplied from overseas through geographically spread sources do not run this risk. The short range of the land based transport system and the consequent limited integration of the food supply, rather than the level of agricultural productivity, increased the incidence of local famine because of normal variation in weather. This risk was severely reduced where towns were connected to multiple regions beyond their contiguous hinterlands.

These cities performed a range of functions. They were centres for political and legal activities. Mass entertainment and religious ceremonies were centred there, although not exclusively. They also performed economic functions. Rome was the recipient of substantial tax and rental revenues but it was not a simple agglomeration of pure consumers.¹⁵⁰ The proportion of the city's population that was engaged in pure consumption included the elite recipients of tax and rental revenues but not to the extent that they were involved in economically necessary functions such as government and administration or in the management of their estates. It included those in receipt of the grain distribution and those dependent on others - again only to the extent that they were not economically active.

The level of urbanisation seen in the imperial Roman economy was not matched until the middle of the eighteenth century. The level of economic activity and the overall per capita production of goods and services that followed simply from the level of urbanisation, likely level of urban productivity compared to rural productivity and the need for large sections of the urban population to be economically active similarly led to higher level of per capita output. These levels were substantially above what was experienced in medieval England, for example. The size of the towns and cities that were possible and which supported these levels of activities required the provision of infrastructure

¹⁴⁸ Removing the assumption that about half the agricultural units were small holdings producing little surplus.

¹⁴⁹ The size of the hinterland increases with the square of the radius.

¹⁵⁰ The view of Rome as a consumer city is a significant oversimplification of its economic structure. Contra Bücher (1906) 371 who considered that the towns of Greece and Rome were pure centres of consumption.

which, as will be discussed later in the chapter on State Finance, depended on state and municipal intervention for its establishment and maintenance.

CHAPTER 4: The extent of trade and why it changed over time

4.1 Introduction

In the last chapter I considered the extent to which agricultural surplus enabled people to live distant from the areas of food production in urban centres with large populations. I also considered the effect that this level of urbanisation had on the level of economic output. In this chapter I consider the connections between these centres of urban population and the economic interaction between those centres. This chapter considers the reasons why the structure that developed during the Roman Empire, of highly populated urban centres spread over a wide area and well connected to each other, led to high levels of economic output. I will concentrate, very largely, on trade which involves sea transport.

4.2 Previous approaches to trade in the economy of the Roman Empire

This subsection sets out a brief summary of the approaches that have been taken by scholars when considering trade in the Roman Empire. There has been much discussion as to the level at which trade grew or did not grow, and the debate is as much informed by social concepts as it is by the incorporation of hard archaeological data. The approach that I will take attempts to broaden the discussion to include explicitly consideration of other aspects such as infrastructure – and to examine trade in the economy of the Roman Empire in such a way that the level of trading activity can be broadly quantified.

Finley, writing in 1973 saw an aristocratic aversion to involvement in commercial activity as a limiting factor. He did not believe that this status consideration was the only factor in play and did not feel able to quantify the relative effect of this status consideration.¹⁵¹ A.H.M. Jones writing in 1974 took the view that trade was ‘relatively insignificant’ because of the high cost of land transport and the fact that markets were ‘very limited’.¹⁵²

The minimalist approach taken by A.H.M. Jones and Finley was opposed, to some extent, in 1980 by Hopkins who proposed a view that trade, along with monetization, had been increasing through the period from the second century BC and stagnated after 200 AD. He attributed this growth to the introduction of monetary taxation which forced the tax-exporting provinces to engage in trade to generate the money required to pay the taxes imposed on them. The introduction of monetary taxation led to movement in goods from these tax-

¹⁵¹ Finley (1973:1999 edition) 60-1. But see Mayer (2012) for an analysis of a ‘middle class’ with different status considerations and which was involved in trade.

¹⁵² A.H.M. Jones (1974) 37-8.

exporting regions to the tax-importing regions, principally the city of Rome and the peripheral provinces in which the army was stationed. The increase in long distance trade was associated with increased productivity and specialization as well as changes in patterns of consumption.¹⁵³ This approach privileges the state as the generator of economic growth and differs from the approach taken by Rostovtzeff, for example, who privileged the individual and his desire for wealth.

As part of Hopkins' evidence for the changing pattern of trade he used shipwreck data to support a conclusion of rapid growth in the last two centuries BC and a decline after AD 200.¹⁵⁴ As I will describe later, and as Hopkins recognised, the use of shipwreck data to quantify trade levels is problematic. Other measures besides the numbers of shipwrecks detected have been used in an attempt to measure changes in the level of trade, such as amounts of amphorae, pottery and ceramic assemblages.¹⁵⁵ Later in this chapter I will present an alternative way to think about changes in trade levels and introduce a measure based on trade infrastructure rather than on numbers of ships wrecked or goods transported.

The approach set out by Hopkins was further developed by Bang, in 2007, who proposed a model based on three concepts: tributary empire, portfolio diversification and protection payment. Most of the economic activity in the empire, he argued, was peasant production for their own consumption. The balance is to be explained as rent and taxes exacted as a protection payment by the elite which in order to monetize the economic value extracted diversifies its portfolio to include other activities such as finance or trade. Trade in this view arises simply to enable an already wealthy land-owning elite to monetize its wealth and is not a wealth-creating activity.¹⁵⁶ In this approach the landed elite and the state are the economic drivers. Rathbone in 2003, based on the evidence of trade related contracts, took the view that there existed an entrepreneurial elite which was interconnected and involved in a range of business activities. In this view class and landed wealth are not explanatory factors.¹⁵⁷ Horden and Purcell emphasis variability in agricultural output because of weather variations and ease of transport across the Mediterranean as factors contributing to trade.¹⁵⁸

¹⁵³ Hopkins (1980).

¹⁵⁴ Hopkins (1980) 105-6.

¹⁵⁵ Wilson (2009b) 229-44.

¹⁵⁶ Bang (2007).

¹⁵⁷ Rathbone (2003).

¹⁵⁸ Horden and Purcell (2002).

Silver, writing in 2008, stressed that the approach taken by Hopkins was a heuristic model, a set of assumptions rather than facts. He argued that the imposition of taxes was likely to reduce production rather than to encourage trade.¹⁵⁹ The conclusion that money taxation drove trade is not a conclusion which is forced by any facts particular to the economy of the Roman Empire. He pointed out, moreover, that in attributing all of the growth to money taxation Hopkins sets to zero the contribution to trade that flows from the desire of individuals to make money. Silver also argued that the picture of an economy based on subsistence peasants was inaccurate, noting the wide range of occupations and that there was specialized production, for example, of oil and wine. In addition the sale of produce by peasants, and therefore trade, had long been unremarkable.¹⁶⁰

Temin, in 2004, followed Hopkins in placing substantial reliance on shipwreck data and using it to support the conclusion that trade reached its maximum in the early Roman Empire.¹⁶¹ Kehoe in 2007 also took the view that there was limited economic growth in the early Roman Empire.¹⁶² Lo Cascio, again in 2007, indicated that many scholars see the first two centuries of the Roman Empire as a period of low growth in GDP.¹⁶³ These low growth views are opposed by Rathbone, in 2003, who took the position that levels of commerce in the Roman Empire were 'enormous and unprecedented'.¹⁶⁴ Robinson and Wilson, in 2011, took the view that conditions were favourable and that there was significant economic growth in the Roman Empire.¹⁶⁵ Harris, writing in 2011, took the position that Roman trade was substantial.¹⁶⁶

Wilson, writing in 2009, attempts to chart the growth of trade. He considers shipwrecks, commodities traded and the containers in which goods were shipped. Each of these attempts faces difficulties which are not resolved. In some cases, such as the case of millstones, there are large amounts of data but there is still an enormous work of synthesising to be done. The difficulties of using shipwreck data to form views on overall levels of trade increase are discussed later in this chapter. Arguments about the general level of trade based on studies of amphorae finds are difficult because of regional changes in consumption patterns and no

¹⁵⁹ I do not find this point convincing although I agree that assuming taxation stimulates trade needs more support than Hopkins gives.

¹⁶⁰ Silver (2008). See also Shaw (2008).

¹⁶¹ Temin (2004) 729

¹⁶² Kehoe (2007) 568.

¹⁶³ Lo Cascio (2007) 619.

¹⁶⁴ Rathbone (2003) 225.

¹⁶⁵ Robinson and Wilson (2011) 1-9.

¹⁶⁶ Harris (2011) 185

clear overall conclusions emerge from the analysis of pottery finds. Wilson argues for an examination of infrastructure as a way to overcome these difficulties. The infrastructure that he focusses on is that required for the production of salted fish and he discerns overall patterns of trade in that commodity. The results of this infrastructure review are broadly consistent with the analysis which I develop in this chapter. Wilson, however, concludes that although there are great quantities of archaeological data it has been difficult to synthesise the material.¹⁶⁷ Approaches to quantifying trade which consider consumption patterns within settlements have been proposed, although these have not been attempted.¹⁶⁸

My approach to the question of quantifying overall trade volumes is not to focus on individual categories of items traded but on infrastructure. The essential feature of this analytic step is to find a parameter which changes with the overall level of trade and which does not vary much for any other reason. As demonstrated later in this chapter, this allows overall conclusions to be drawn as to when and at what rate total trade volumes increased and declined and the institutions which contributed to those changes can be identified. The conclusions can also be set in geographical contexts, because trade did not increase and decline everywhere all at once. I believe that when clarity has been secured in this area then the important questions of what was happening at the individual commodity level can be answered with more confidence. A decline in the export of a commodity at a time when overall activity is expected to be increasing then raises questions as to changes in patterns of consumption and production, changes in comparative advantage, the emergence of competition or changes in patterns of competition and of substitution and there is no need to use that decline as indicative of overall trade decline. The parameter which I use is the length of the land-sea interface, that is quays.

I take the approach that trade growth increases substantially through the first and the beginning of the second century AD and was at a different order of magnitude at the end of that period compared to what it had been at the beginning. The first century AD saw the onset of a trade boom. I do not see, moreover, the question as being simply one of growth or no growth but of where growth occurred and where it did not. The growth of trade in Africa and in Rome was different than that in Egypt. I present a template for such an alternative

¹⁶⁷ Wilson (2009).

¹⁶⁸ Harris (2009) notes the difficulties associated with shipwreck data. The indications are inconsistent with textual and land site data. He suggests extending the analysis to commodities which leave no trace in the archaeological evidence and to the use of settlement data to generalise consumption patterns. Fulford (2009) recommends analyses based on the different patterns of consumption in the settlement hierarchy.

quantification later in this section. Unlike Finley I do not see an elite aversion to commerce as a hindrance to the development of trade. On the contrary, in this social economic structure, non-elite members are agents for substantial growth in trade. I further disagree with Finley in that I believe that elite members did play a role in trade and there was a great variety of ways in which the elite could involve themselves in this activity. I disagree with Hopkins and Bang in their view that the state – and the landowning elite – were the drivers of demand; the state had a role in the development of trade but it was the (often accidental) one of facilitation rather than being the driver of demand.

4.3 Trade and Theory

This section discusses the economics of trade in a theoretical perspective. The purpose is to provide a theoretical framework that will help us first to identify the economic factors that made trade possible, and second to propose what trade may or should have done for the economy. An underlying assumption is that from the diversity of economic theories which are appropriate for the very wide range of modern economies – which include those of Saudi Arabia, the Soviet Union, North Korea, Somalia and the United States – a coherent framework can be generated for the imperial Roman economy. The intention is to apply a theoretical framework which enables evidence on infrastructure capabilities, trading structures and social attitudes to be used in the quantification of levels of trade and changes in levels of trade.

4.3.1 Why goods are moved

The movement of goods from one location to another is classically explained by differences in the price at which the same good is exchanged in different regions. In regions A and B the price of good G is formed by local supply and demand, which generates two prices – $P(A, G)$ the price at which the good exchanges in region A, and $P(B, G)$ the price at which the same good exchanges in region B. If the conditions for trade are met then goods are exchanged between the two regions and classically a new price emerges which is intermediate between these two.¹⁶⁹ If initially $P(B, G)$ is higher than $P(A, G)$ then goods will be traded from region A to region B. Production of that good will increase in region A and fall in region B; consumption will conversely fall in region A and increase in region B. More of the good will

¹⁶⁹ In practice there are frictional costs in moving goods between the two regions and other imperfections and rather than the two prices converging to one price they converge but regional differences remain.

be produced in the region which can do so at the cheaper price and that excess will be exported. Trade encourages specialization and tends to increase productivity.¹⁷⁰

The classic explanations for the benefit of trade, and a partial explanation for the forces which stimulate it, also rely on consideration of the comparative advantage that the trading regions have in the production of different goods. If the marginal cost of producing good G in region A – as measured in terms of the number of units of another good H which might alternatively be produced – is less than the marginal cost of producing good G in region B, then region A has a comparative advantage in the production of that good G. So, for example, if in region A land and labour can be applied to produce either 3 kg of wheat or 2 litres of wine whereas in region B the same land and labour produce 2 kg of wheat or 3 litres of wine then A has an advantage in the production of wheat and B has an advantage in the production of wine.¹⁷¹

In the autarkic state both goods G and H are produced in region A. In this example, in the autarkic state, region A produces wine even though it is relatively more productive in growing grain. Where trade takes place between the two regions more of good G is produced in region A and less of good H. Part of the higher production of good G is traded for good H produced in region B and part is available for consumption in region A. The effect of trade therefore is for regions to specialize in the production of those goods in which they have comparative advantage and for there to be growth in the aggregate amounts of goods produced. Region A tends to concentrate on wheat production and region B on wine production. The quantity of goods produced increases when there is trade between the regions.

Comparative advantage can arise because of differences in climate. Differences in the relative availability of the factors of production – land, labour, physical and human capital – can also produce comparative advantage and disadvantage. A region which is very well supplied with a particular factor of production will have comparative advantage in the production of goods which are intensive in that factor. So, for example, a region which is relatively well endowed with low-skilled labour will have a comparative advantage in the production of products which are intensive in low-skilled labour. A region may have an absolute advantage where it

¹⁷⁰ See Krugman and Wells (2013). The amount exported will also include the amounts which the consumers in region A can no longer afford because of the increased price.

¹⁷¹ In this case the regions have absolute advantages. If in region B the same land and labour produces 1 kg of wheat and 1 litre of wine then B has a comparative advantage in the production of wine.

is endowed with a factor of production that another region completely lacks – for example where it has gold mines that its trading partner does not have. Comparative advantage can also arise from differences in technology. Such an advantage could be where one region has the technology required to produce glass and the region with which it trades does not.¹⁷²

4.3.2 Risks and costs which act as deterrents to trade

The venture only happens if the instigator of the venture believes that the positive possibilities sufficiently outweigh the negative possibilities. The more variability there is, the less likely it is that the venture will be undertaken, or the less likely than it will be financed by debt. Also the more variability there is, then the higher the positive outcomes need to be for the venture to be undertaken. In other words the more variability or risk that exists in the trading operation, the fewer the number of ventures undertaken and the higher the returns to those who undertake them. This section attempts to place the trading risks that surrounds these ventures into context. Risks are factors which contribute to the variability of $S(G, B)$.

The process of moving goods from A to B involves risk. If, for example, the process involves transport by sea then the trade risks include the risk that the ship will sink. To an extent this risk could be mitigated by maintenance of vessels and by choice of sailing times. More significant than the risk of sinking is possibly the risk that the vessel is blown off course or becalmed. If it does not reach the intended destination or arrives late, then goods might not be sold or contractual arrangements not met. Piracy can be a significant risk. The level of maritime predation may however be significant, not so much for its impact on the profitability of ventures as for the increased likelihood of death faced by traders travelling with their goods. Inflation presents a risk to trade by introducing a reluctance to use money. In times of hyper-inflation there is general desire to dispose of money and a reluctance to exchange goods for money. However, the most significant risk can be commercial. Shippers face the risk that there is no cargo for them to ship, whether because of an oversupply of shipping capacity, port closure, adverse weather or downturn in demand, for example. The most significant risk for traders can be that the price at the destination is such that they cannot sell their goods at a good price or at all. Shippers are incentivised to find cargo and traders are incentivised to find information. These risks are partly managed at ports by information exchange between shippers and traders.

¹⁷² See Krugman and Wells (2013) 212-221 for a standard explanation of comparative advantage and illustrative examples.

4.3.3 Intermediation

The movement of goods between regions can happen because producers in region A deal directly with consumers in region B or through the intermediation of traders. The trading business is an adjunct to the production business. The profit and the risks must provide enough incentive to move goods from one place to the other.

The opportunity for profit for the trading venture is:¹⁷³

Profit (G, A, B) = S (G, B). (1-T1) –O (G, A). (1-T2) > Profit requirement Formula 4.1

Where S(G,B) is the sale proceeds achieved for the quantity of goods G at location B, O(G, A) is the corresponding outlay at location A and E1 and E2 are the transaction costs incurred at those two locations – expressed as proportions of the value of the goods at the location.¹⁷⁴

The sale proceeds depend on the likelihood of the cargo reaching its destination, the proportion of the cargo which is actually sold and the price that is achieved. Profit (G, A, B) is the profit from the venture of transporting goods G from location A to location B. This profit is available to those who have financed the venture – whether through risk capital or through loan capital. Each trading venture is subject to risks – the risks associated with trade in this period are discussed later in this section – and Profit (G, A, B) is a variable whose value is unknown at the start of the venture.

4.3.4 The prerequisites for trade

There are some prerequisites for trade, and changes in these factors affect the level of trade. These prerequisites include that it is possible to move good G between locations A and B, that there is a demand for good G at location B and that a suitable price can be offered there. Information flow is critical to the volumes of trade. Ventures will only be undertaken when traders at location A have a means of knowing what prices are – or might be - at location B with sufficient confidence to undertake the venture. The level of demand at location B also needs to be known when assessing quantities of goods to ship.

¹⁷³ Note that what we are looking at here is the profit of the trading venture. The shipping business is a different matter – and has very different risk/return characteristics.

¹⁷⁴ I only include as transaction costs such items as freight charges and customs dues which are known. I do not include as expense risk factors such as venture failure because of predation. Nor do I include financing costs, which are negotiated allocation of the proceeds between financier and merchant because I am considering the total proceeds of the venture.

Although trade can happen through the exchange of goods by barter, the availability of money significantly increases the level of trade. Prices are more transparent, there is greater flexibility in exchange of goods and there is greater variety in the ways in which investors can participate in the ventures

Another prerequisite of trade is that at location B there be suitable structures for the reception of goods, their storage and onward transmission to markets or direct to purchasers. To the extent that these structures do not exist, then flows will not happen. Also to the extent that there are improvements to these structures, then trade will increase.

4.3.5 Scale of profits

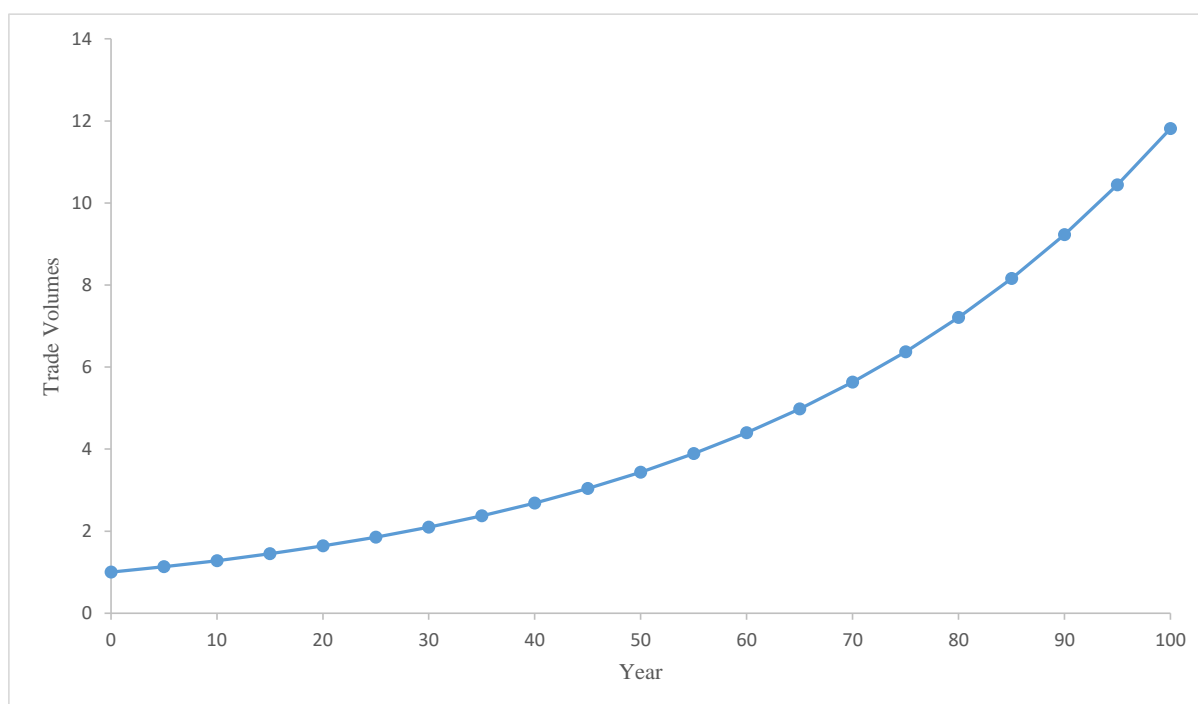
The assets of these operations are cargo and cash. Cargo is being moved continually, converted into cash which is then converted into cargo. The extent to which cash is reinvested into cargo depends on perception as to the potential profit and risk in the available trading opportunities. It also depends on the extent to which traders retain the profits generated from ventures or extract and consume the proceeds. The rate at which traders choose to withdraw from the trading business affects the total size of the trading enterprise. Cash may also be withdrawn to pay down loans advanced. The level of activity is also dependent on the attractiveness of trade for new investment and new finance. A trading business which is retaining the excess cash generated by profitable ventures needs to identify outlets for that excess cash. These outlets might be either extension of existing opportunities – buying more of the same goods – or come from the identification of new trading opportunities. Limiting constraints on reinvestment include limitations on demand for goods already being traded or the inability to identify new opportunities.

Trading structures can place constraints on the extent to which profits can be retained. For example, sole trader operations which are owner-managed have growth rates which are constrained by the scarce resource of the owner's time and have limited growth rates and therefore retained earnings. Modern structures such as joint stock companies have separate owners and managers and can permit higher growth rates. The *societas* structure, devised in the first century BC, and briefly described later, is one such structure.

The basic trading model is that goods are acquired at location A, transported to location B and then sold or otherwise exchanged for consideration which exceeds the outlay. If part of that excess is retained within the trading business then the value of the goods transported increases. In conditions where trade is favoured and where earnings are retained, substantial

growth can be achieved. Where, for example, there is a 10% return on trading activities undertaken in a year and a retention rate of 25% of the profits in the business, then aggregate trading business will increase by more than 10 times in the space of 100 years. As will be discussed later this level of return is probably much less than was achieved in the period under consideration. The level of reinvestment is also low relative to what was possible. Thus very substantial growth is theoretically likely in peaceful times with a stable currency and a group of entrepreneurs reinvesting profits from trade. The following graph illustrates this level of growth in values of goods transported on these assumptions. A long period of stability enables this compounding effect to occur. What is also needed are opportunities for the retained earnings to be profitably deployed. These opportunities are discussed later in Chapter 6 on knowledge.

Figure 4.1 Schematic illustration of the effect of profit retention on trade



A stable political environment, access to wide area subject to the same legal system, elimination of risks such as predation, currency certainty and adequate transportation infrastructure, for example, are factors which favour profit retention and therefore encourage trade growth. If these conditions exist for a reasonable period of time then substantial growth in volumes of trade may occur.

4.4 What goods were moved and why

This subsection provides a description of the processes of production, comparative advantage and changes in trading patterns for a sample of goods traded in the Roman Empire. As the political structure of the Roman Empire changed and as knowledge spread, there were shifts in the geographic distribution of comparative advantage and in trade patterns.

Most of the goods being traded were items of everyday consumption and use and were in demand by non-elite members of society as well as the elite, although there were differing levels of demand for high quality goods and produce. Significant growth in trade implies that there has been increased consumption by the non-elite, more or better quality food, or more or better quality tableware, for example.¹⁷⁵

4.4.1 Wheat

The production of grain is intensive in the factor of land and is less intensive in labour. The labour required is largely low-skilled and the basic technology and knowledge required was long established and widespread, although there were undoubtedly differences in the application of this technology. Capital required was largely seed, usually saved from the previous year's production, and for implements which were relatively basic and could be produced locally. Comparative, and in some cases absolute, advantage derived from climate. Weather conditions at any one location are variable year on year and the comparative advantage derived from climate that one location has over another also varies. Egypt, which had the most favourable conditions, had some of the lowest wheat prices. Those in Sicily were some 40% higher and in Italy prices were about 400% times the level seen in Egypt. Prices along the eastern Mediterranean coastline may have been comparable to Egyptian prices.¹⁷⁶ These prices reflect differences in climate conditions, Egypt for example benefitting from the annual inundation of the Nile, and with land naturally fertilized by the Nile rather than the relative availability of labour or capital.¹⁷⁷ They also reflect the differing balance between the local supply and demand and point to significant trading opportunities. These trading opportunities vary from those which are perennial, for example between Egypt and

¹⁷⁵ If the elite represents 10% of the population, say, then a large increase in demand, 50% say, is an overall small increase of 5%. A large increase in demand by the non-elite of 50% leads to an overall large increase of 45%. Large increases in trade are to be associated with a general improvement in the standards of living of the non-elite.

¹⁷⁶ See Rathbone (2009a) for a discussion of these price levels.

¹⁷⁷ Kessler and Temin (2005), quoted in Rathbone (2009), argue that prices are to be explained as a single Roman price adjusted for transport costs.

Italy where the price difference is so high that substantial profits are always available, to the opportunities which arise temporarily because of shortages caused by adverse climate variations in a region. Political changes may also have had effects. So, for example, the acquisition of Egypt and the freeing of trade between it and Rome may have increased the importance of the comparative advantage that it had in the production of grain leading to increased Egyptian output and the diversion of Italian land from the production of grain.

4.4.2 Wine

The production of wine again depends on climate and is intensive both in the factor land and in capital. Variations in weather conditions can have significant effects on the size and quality of the harvest. The year on year price variation noted for wheat is also observed for this foodstuff and short term trading opportunities arose as adverse weather conditions led to wine scarcity.¹⁷⁸ Price variations also occur because of variations in the quality of the product. Changes in patterns of demand for wine affected trading patterns. Imports of wine into Gaul reduced from the middle of the first century BC and into Britain from the end of the first century BC.¹⁷⁹ This decline is partly explained by a reduction in local consumption, at least in the case of Britain, but also by an increase in Italian consumption. Exports from Spain also substituted for Italian exports.¹⁸⁰ Northeast Spain was a source of imports to Italy from the first century AD.¹⁸¹ As the city of Rome grew there was considerable expansion in the demand for wine. Wine is, however, a highly differentiated product and competition patterns for high and low quality wine can differ. It is possible that domestic production of wine was increasingly at the lower end of the market.¹⁸² This pattern of shifting comparative advantage occurred in the production of other commodities.¹⁸³

¹⁷⁸ See Chapter 7 on the money system.

¹⁷⁹ Tchernia (1983) describes the wide diffusion across Britain and Gaul of amphorae used to transport Tyrrhenian wine in the second and first centuries BC.

¹⁸⁰ Sealy (2009) 25-6.

¹⁸¹ Parker (1990) 343.

¹⁸² Purcell (1985) but see Tchernia (1988) for the view that Italian production was high quality for local consumption, supplemented with imports from Gaul (and Spain) of cheap wine.

¹⁸³ This pattern of a reduction in Italian production and an increase in overseas production can be seen also in the cases of olive oil and terra sigillata pottery. Knowledge of how to cultivate olive oil had spread as the provinces became integrated as with the migration of Italian skilled labour to Spain which became in the first three centuries AD the largest producer and exporter of olive oil in the Roman Empire. West (1929). (Spain currently produces over 40% of world consumption of olive oil.) This increase in production was accompanied by a decline in Italian exports. Greene (1986) describes the continual decline from the first century BC of Italian amphorae found at Ostia. These number just under 80% at the beginning of the first century BC and by the mid-second century AD had fallen to negligible levels. This pattern can also be seen in the production of terra sigillata pottery. The export of Italian ware declined from the early part of the first century AD being replaced by the end of that century by production centres which had more local distribution, for example Southern Gaul

4.4.3 Overall patterns

There is evidence that there were significant shifts in patterns of trade through the time of the Roman Empire. There is, for example, a reduction in the importance of Italy as an export area for agricultural produce. These changes were occurring as technology spread and as various new territories became integrated into the system. There is no evidence that any of these changes in trading patterns were being planned but rather that they are happening because of changes in the balance of comparative advantage.

4.5 Risks and costs that restricted trade

This section summarises and gives a broad quantification of the risks and costs of trade in the Roman Empire.

4.5.1 Shipwreck and piracy

The risk of the ship sinking on a venture was low relative to the other risks to the profitability of the venture. It was probably about 1% and possibly significantly less than this.¹⁸⁴ Hopkins in 1980 considered shipwreck data in the western Mediterranean as a proxy measure for trading activity.¹⁸⁵ The results are summarised in the following table.

and Germany, or by African slip-ware from Tunisia which served the Mediterranean. While production continued in Italy its quality declined and was replaced by exports from, for example, Southern Gaul. The change in the location of production may have followed from the dispersal of the knowledge required.

¹⁸⁴ If it had been as high as 10% for example then a ship which took two round trips a year would be expected to have an average working life of about three years and to have almost certainly sunk within six years. At this level of risk total freight charges would hardly have covered the cost of replacing vessels that sank. The useful life of a ship likely extended into the decades. Ships were leased for five to six decades – see Rathbone (2003) 202, for example. The probability of sinking was possibly comparable to that experienced now in Indonesia where wooden vessels are used. Sinking rates there appear to be less than 1% per annum – or significantly less than 1% per venture. Discussion with Phil Skelton.

¹⁸⁵ Hopkins (1980) 105-6 using data from Parker (1980). Parker (2007) 187 shows a similar pattern of a peak at the end of the Republic and start of the Empire.

Table 4.1: Shipwreck data – used as a proxy for trade

1: Period ending	2: Number of shipwrecks	3: Implied annual trade growth in trade
400 BC	25	
200 BC	50	0.3%
AD 1	160	0.6%
AD 200	140	0.0%
AD 400	80	-0.3%
AD 650	30	-0.2%

The shipwreck data, however, in Hopkins' collation of the numbers, indicates a decline from the beginning of the first century AD. He attributed this feature of the data to a reduction in the level of piracy and warfare in that period and to chance variations in the data. In other words the shipwreck data in Table 4.2 does not tell us what happened to levels of trade between the first and second centuries AD.¹⁸⁶ The conclusion that Hopkins then drew from the data was that in the period 200 BC to AD 200 there was 'more sea-borne trade in the Mediterranean than ever before, and more than there was for the next thousand years.'

The use of this data as a proxy for trading activity has been criticized. The data is in any case only weak evidence for trading activity. The numbers of ships wrecked depends on factors which are not necessarily constant over time; indeed shipwrecks could as well be used to measure changes in those factors. Changes in maintenance standards, marginal improvements in ship construction or the early retirement of older vessels will also affect the rate at which ships sink. Changes in types of cargo carried or the containers within which the cargo was transported can also affect the rates at which shipwrecks are detected.¹⁸⁷ Changes in trading patterns, rather than decline in trade volumes, may reduce the volumes which pass through

¹⁸⁶ More recent work on shipwreck data, such as Wilson (2009b) 219-29, also gives no clear indication of trends between the first and second centuries AD. See Horden and Purcell (2000) 368-72 for discussion of shipwreck data and, in particular, a comparison which shows the numbers of shipwrecks in the Roman period peaking at a figure which is about six times the peak seen in the medieval period.

¹⁸⁷ Wilson (2009b) 219-21 notes that shipwreck remains generally survive by being protected by amphorae or marble lying on top of them. Ships carrying other cargoes do not leave remains. The replacing of amphorae with barrels results in changes in detection rates.

the areas explored by marine archaeologists.¹⁸⁸ Ships are also wrecked by piracy and changes in predation rates will affect the number of shipwrecks. The data as presented are for broad periods of time, but when the data are presented for shorter periods, they suggest the conclusion that trade did not follow a simple pattern of constant growth followed by constant decline, which is contrary to expectations.¹⁸⁹ Nevertheless, the use of this data as a proxy continues to be accepted even where there are misgivings.¹⁹⁰ The contrary indications of other data which are stronger evidence for trading patterns are to be preferred. This evidence, which is discussed in the next section, is stronger because change in trading activity is very likely the main or the only reason associated with changes in these variables.

Piracy was a substantially reduced risk in this period and even at the higher levels seen before Pompey in 66 BC reduced the level of maritime predation was probably significant not for its impact on the profitability of ventures as for the increased likelihood of death faced by traders travelling with their goods.¹⁹¹ From the time of Augustus the Mediterranean was a Roman lake policed by an imperial navy and there were no safe bases for pirates. As I will show in this chapter there was a very substantial increase in trade in the first two centuries AD. If this were reflected in the shipwreck data then instead of 140 shipwrecks for the period AD 1 to AD 200 we should have something of the order of 500 to 800 shipwrecks. I believe that the low level of shipwrecks is more likely to be attributable to the reduced rate of predation and possibly to improvements in ship quality or management.

4.5.2. Inflation

As noted earlier, inflation presents a risk to trade as it introduces a reluctance to use money now for possible future benefit and there is a disincentive to sell goods in exchange for money which will be devalued. It also introduces a disincentive for financiers to support ship

¹⁸⁸ Wilson (2009b) 223 describes the changes in trading patterns between southern Gaul and Italy between the late Republic and the first century AD. In the earlier period there was substantial wine trade between Italy and southern Gaul, which was supplied by local sources thereafter. This change in trading patterns reduces the numbers of ships wrecked.

¹⁸⁹ See Tchernia (2011) 161-2.

¹⁹⁰ Shaw (2008) 98-9 uses the shipwreck data as an indication of trading volumes (despite his misgivings) because it is consistent with the variations in the numbers of coins minted, papyrus finds and the levels of air pollution detected in north European lakes and in Greenland. The shipwreck data is not consistent with the indications from the numbers of coins minted, as discussed later. The indications from Greenland show a 'spectacular' increase around the middle of the first century AD (de Callatäy (2005).

¹⁹¹ Scheidel (2009) 36 – 'the comparative evidence in this paper strongly suggests that the cost of predation was a crucial determinant of trading costs'. A 5% risk per voyage that a ship is attacked implies a high level of predatory activity but is low relative to the commercial risks. It is extremely high relative to normal mortality expectations. Per discussion with M. Crawford there is no evidence that piracy was responsible for a significant number of shipwrecks.

construction and the purchase of cargo. These disincentives emerged in the middle of the second century AD with the onset of inflation and became extremely severe in the last quarter of that century.¹⁹² There was, therefore, significant downward pressure on trade through that period.

4.5.3 Commercial risks

The commercial risk that there was insufficient demand for product at the destination point or that prices were inadequate was very likely the dominant risk. The Roman knight Rabirius made a loss when his ships carrying paper, linen and glass from Egypt finally arrived in Puteoli. Many ships were packed with these and the proceeds did not even compensate for transportation costs.

“... We have heard that ships belonging to Postumus arrived at Puteoli, and merchandise was seen there, things only showy and of no real value, made of paper, and linen and glass; there were several ships entirely filled with such articles; but there was also one little ship; the contents of which are not known.” That voyage to Puteoli, (such was the conversation at the time), and the course taken by the crew, and the parade they made, and the fact, too, of the name of Postumus being rather unpopular with some spiteful people on account of some idea or other respecting his money, filled in one summer numbers of ears with these topics of conversation.¹⁹³

4.5.4 Costs

The following table shows a sample of the levels of charge that could be imposed for the transportation of wheat, as set out in Diocletian's Edict.¹⁹⁴ I believe that the freight charges can be best read in the context of the number of round trips that a shipowner might make. In the case of the Alexandria to Rome route he might make one, or possibly two, round trips in a year, and in some cases returning with a cargo consisting of ballast. In the case of the Levant-Spain route he might make an outward trip, and probably a full round trip, in a year. In column (4) I show my assumption as the average number of round trips that could take place

¹⁹² See Chapter 7.

¹⁹³ Cicero *pro Rabirio* 39-40. Loeb translation.

¹⁹⁴ The figures are taken from Arnaud (2007). I also take the view that these charges reflect underlying supply and demand for ships along the routes indicated. I do not read them as being fully explainable by distances as e.g. Arnaud (2007) because much of the freight charge relates to loading and unloading costs. I believe that they are better explained in terms of the number of round trips a ship might make in a year.

on each route. I then show in column (5) the estimated ship-owners annual income, expressed as a percentage of the value of his ship, before costs.¹⁹⁵

Table 4.2: Freight charges for wheat: based on Diocletian's Edict

1: Origin	2: Destination	3: Freight charges (percentage of value of cargo)	4: Number of round trips	5: Ship-owner's annual income (percentage of value of ship)
Alexandria	Rome	16%	1	16%
Alexandria	Africa	10%	2	20%
Alexandria	Sicily	10%	2	10%
Levant	North Africa	16%	1	16%
Levant	Spain	20%	1	20%
Turkey	North Africa	8%	2	16%
North Africa	Sicily	6%	3	18%
Sicily	Marseilles	8%	2	16%
Rome	Spain	10%	2	20%
Levant	Portugal	26%	1	26%

These freight charges indicate that marine transportation was as cost efficient in the Roman Empire as it is in modern times, where unloading is manual rather than by crane lifted containers.¹⁹⁶ The annual income figures, shown in column (5), indicate that the shipowner's annual revenue might be between 15% and 20% of the value of his ship, less the costs of

¹⁹⁵ The ratio of the value of the ship to the value of the cargo is a key parameter affecting the profitability of shipping. For grain cargo I have used a ratio of 2. Rathbone (2003) 202 uses an average construction cost of a ship of between 1.25 and 1.5 times the value of the wheat it could carry and Hopkins (1983) 101 shows a ratio of construction cost to wheat cargo of 1.4 to 2.2.. Ratios for higher value cargos are lower and the transportation is consequently more profitable. There was a financial incentive to shippers to replace as many sacks of grain as they could with more valuable and less bulky cargo. I believe that this the main reason that the state authorities felt impelled to offer inducements to shippers to transport the *annona*.

¹⁹⁶ Conversation with P. Skelton. Comparison is with modern small open vessels travelling from the Ukraine to the Mediterranean rather than containership. These charges are broadly comparable to seventeenth century European prices, see Scheidel (2011) 35.

loading and unloading, which could be high.¹⁹⁷ The implied returns on capital, after costs are deducted, may not have been far distant from the risk free returns.

The freight charges, shown in column (3), are low relative to the difference in the prices of wheat at the different locations, 40% between Egypt and Sicily and 400% between Egypt and Italy and indicate that in the distribution of profits the greater part accrued to the trader. As will be discussed in the next subsection, barriers to entry into the shipping business were low, there was presumably no shortage of ships and shipowners' bargaining power was consequently limited. Other costs needed to be deducted such as customs duties, which added 25% for import into the Empire from the east and 2.5% into and out of custom zones.

These freight charges also provide further support to the view that piracy and shipwreck risks were low. The profit to the shipowner was these charges less his costs, including the costs of loading and unloading cargo, which could be substantial, and if piracy and shipwreck had been anything more than 1% per voyage then this business would have been uneconomic.

Overall, through the period from the beginning of the first century AD until well into the third century AD the risks and costs that might restrict trade were low and there was nothing to prevent the growth of trade. The normal commercial risks to traders that they might lose money because of lower than expected demand or prices for their goods remained throughout the period. Towards the middle of the third century AD and with increasing force from the middle of the second half of that century the onset of inflation and then hyperinflation acted as a dampener to trade.

4.6 Intermediation

4.6.1 The Ship-owner

Owners of ships, who were normally men of 'modest wealth and status', usually had just one vessel and evidence for multiple ownership is rare.¹⁹⁸ Ship-owners did, however, at least sometimes form themselves into *collegia*, legally constituted associations of individuals.¹⁹⁹ They had the elements of mutual insurance providing funeral coverage, were social groups holding banquets and found civic expression in processions through the towns. Socio-

¹⁹⁷ Rickman (1980) 23 notes the large labour forces required. See Russell (2009) 116-7 for comments on the costs of transporting stone objects, rather than ceramics. Most of modern freight charges are to cover the costs of transferring goods from the docks onto ships and from ships to the docks. Conversation P. Skelton.

¹⁹⁸ Rathbone (2003) 204.

¹⁹⁹ van Nijf (1997) 243 believes that these associations typically had a membership which consisted of men of middle rank although Wilson (1996) considers that membership was mainly drawn from the lower strata.

political linkages to the local elite were formed through the inclusion of benefactors, who might fund banquets, and patrons, who might be local magistrates and who could be looked to support the interests of the membership.²⁰⁰

Ship-owners could participate in trade through different sized businesses. Most ships sailing inter-provincial routes were between 60 and 80 tonnes burden, although larger ships were sometimes used on these routes. The very large ships of size 200 to 400 tonnes burden sailed on the main shipping routes – such as from Alexandria to Rome – and across the ocean from India to the Red Sea ports. These very large ships were also used for the transport of heavy goods such as marble.²⁰¹ By way of comparison the ships used by the East India Company at the very beginning of the eighteenth century were around 300 tons and by the middle of the century had increased to about 500 tons.²⁰² There is little evidence of technological improvement in vessels through this period. Indeed in the period from the first century BC to about the middle of the nineteenth century there is no appreciable change in the range of ship size.²⁰³ Some improvements in container technology did occur.²⁰⁴ For example there was an increase in the use of barrels rather than amphorae to move wine. This development improved the economics of trade as far as the margins of merchants were concerned, since it increased the carrying capacity of ships. Barriers to entry were low. Knowledge of how to construct ships was widespread and very long established, the materials were relatively easy to come by and the even the basic capital required could be borrowed. Construction was therefore easy and cheap. As noted earlier, profits from this business were low.

4.6.2 Merchants

Merchants similarly carried out their business as individuals, although partnerships were common.²⁰⁵ In some cases merchants travelled with their cargo and the need to do this may have been a significant limitation to business growth. Associations of merchants by type of good traded are evident in the Piazzale delle Corporazioni in Ostia where we see groupings of

²⁰⁰ See Wilson (1996) 3,10 and van Nijf (1997) 243-7.

²⁰¹ Rathbone (2003) 210.

²⁰² Keay (1993) 220.

²⁰³ Parker (1990) 340 notes that the size of a ship wrecked off the coast of Albenga in the first century BC was between 500 and possibly 600 tonnes burden. Parker (1990) 341 indicates no appreciable change in the length of ships over the period 160 BC to AD 420. Houston (1988) 554-6 indicates that in the sixteenth century scarcely any ships docking at London were of more than 200 tonnes. Lloyds Register indicates that by 1812 ship size had increased but few ships were of a size greater than 400 tonnes. The picture changed by the start of the twentieth century when no ships in the British merchant navy were of less than 1,000 tonnes burden.

²⁰⁴ Although not in ship construction. Hopkins (1983) 97 notes that there were no improvements in ship construction in the period 200 BC to AD 400.

²⁰⁵ Harris (2011) 181.

traders in rope, leather, wood and grain. These associations need not, however, indicate that the individuals were formally incorporated into a *collegium*. Activity overlapped between the two categories of owner of the vessel and owner of its cargo and in some cases the ship-owner was engaged in the business of transporting his own cargo.

In contrast to the business of owning a ship, fortunes could be made through trade. The Muziris papyrus documents a venture from Alexandria to India and back, carrying on the return journey a mixed cargo, including ivory and nard, in the mid second century AD. The shipment was sold for HS 7 million at Alexandria and the prices achieved there, after the 25% tax, were probably at least four times greater than what the trader had borrowed to finance his venture.²⁰⁶ If the trader had borrowed HS 1.5 million then his profit was something in the order of HS 5 million (less the costs such as that of hiring the ocean ship, the trans-desert transport and the Nile barges) and the profit to the financier was about HS 0.5 millions.²⁰⁷ What presumably is happening here is that the trader has the strategic assets of knowledge of where to source the materials, overseas local contacts and the ability to manage the journey. The financier either lacks these assets or chooses, as Trimalchio suggests – see below – was the norm for a man of great wealth, to concentrate on lending rather than on commerce. The amounts of profit available to the trader are substantial, in this case more than the minimum senatorial census. The rewards of trade, which the trader kept, provided strong motivation for a man at any social level to capitalise on his knowledge and abilities. Through trade a man of limited means might acquire a fortune not available to him through labour and sufficient to place him alongside the elite, even if it did not give him their social standing. I contrast this with the approach taken by Finley ‘... new wealth came from war and politics (including such by-products as tax farming), not from enterprise.’²⁰⁸

The story of Trimalchio further illustrates the position of the individual. Underneath this story there are some common assumptions which are made by the meta-narrator Petronius and presumably accepted by his elite audience. Trimalchio is fabulously wealthy. What is in conflict is whether he did in fact make this money through trade, as he claims or through inheritance as the narrator pushes us to believe. The story was written in the time of Nero which is when the trade boom was starting, as we shall see, and the fabulous wealth accruing to those who did not disdain trade, and had the capabilities needed to profit from it, may

²⁰⁶ Rathbone (2000) 49. Note it seems that the financier buys the cargo from the trader.

²⁰⁷ Using an illustrative 20% interest rate and assuming that the financier lent all the capital.

²⁰⁸ Finley (1965) 39.

have been a sore point. The essential point, however, for our purposes is that a considerable fortune can be made from trade.²⁰⁹ Explicitly, the underlying clear common assumptions in this story are that

- Trade is an entry point to great wealth.
- Individuals of non-elite standing can access that wealth.
- The state allows those individuals to keep the wealth they have generated.

The requirements for trade as described by Trimalchio are the ability to identify an opportunity, access to finance and to have good luck and staying power. We should also note that he describes the progression, implicitly usual, as being a movement from commerce into finance.

4.6.3 Trading Structures

Ship owners and merchants could form business associations, which could additionally include financiers. The *societas* had evolved as a legal structure during the Republic, and enabled the different participants in a business enterprise to share costs, capital, individual effort and expertise and profits and was a form of partnership. A *societas* could be formed for a variety of reasons including to collect indirect taxes, to provide military supplies under state contract, to lend money, to undertake urban waste disposal and for trading purposes, for example. The association could cover single or multiple ventures and could be in existence for a number of years and some had management structures with executive officers and record-keeping.²¹⁰ Plutarch describes the formation of one such *societas* by Cato. Rather than investing directly in shipping and trading, Cato formed an association of fifty *socii*, shippers and traders, and lent money to that association through a freedman who was also a *socius*, presumably the balance of the capital came from the *socii* or external financiers.²¹¹ This passage has been used to support the position that the elite viewed trade negatively, that consequently trade was limited and that the attitude to trade in the ancient economies was such that these economies were fundamentally different to modern economies.

²⁰⁹ See Weber (1978) 1359 'As a group, the freedmen can be best likened to our petty capitalist middle class (which at times may accumulate considerable wealth) ...' I believe that wealth creating opportunities were available to others including soldiers, who had the opportunity to accumulate savings and knowledge during their service.

²¹⁰ Rathbone and Temin (2008) 235. Rathbone (2003) 214.

²¹¹ Rathbone (2003) 213-214. Plutarch, *Cato Maior* 21.

However, as he applied himself more strenuously to money-getting, he came to regard agriculture as more entertaining than profitable, and invested his capital in business that was safe and sure. He bought ponds, hot springs districts given over to fullers, all of which brought him in large profits, and “could not” to use his own phrase, ‘be ruined by Jupiter’. He used to loan money also in the most disreputable of all ways, namely on ships, and his method was as follows. He required his borrowers to form a large company, and when there were fifty partners and as many ships for his security, he took one share in the company himself, and was represented by Quintio, a freedman of his, who accompanied his clients in all their ventures. In this way his entire security was not imperilled, but only a small part of it, and his profits were large. He used to lend money also to those of his slaves who wished it, and they would buy boys with it, and after training and teaching then for a year, at Cato’s expense, would sell them again. Many of these boys Cato would retain for himself, reckoning to the credit of the slaves the highest price bid for his boy. He tried to incite his son also to such economies, by saying it was not the part of a man, but of a widow woman to lessen his substance. But that surely was too vehement a speech of Cato’s, when he went so far as to say that a man was to be admired and glorified like a god if the fiscal inventory of this property showed that he had added to it more than he had inherited.²¹²

There are two stories being told here. The story that Plutarch, a Greek writer and minor official, tells us about Cato and the story that Cato, a new man and Roman Consul and Censor, tells us about himself. Plutarch injects into his consideration of business ventures a value system which sees lending to trading ventures as disreputable. But it is in only this one case that he is explicitly taking a moral stance. He does not object to the other ways in which wealth is created and has no difficulty with the accretion of wealth beyond the amount inherited. His difficulty in that last case is with the excessive praise of those who make money, which he attributes to Cato. Cato clearly has no problem with making money in any of the ways described and has clear opinions on the best ways to do it. There are two different value systems at play here and no evidence that the elite attitude to trade was any different to now. Then as now different people looked at the same thing differently and did different things.

²¹² Loeb translation.

It is also not obvious that Plutarch fully understands Cato's trading ventures since it is difficult to see how Quintio could accompany all fifty ships unless they were forced to sail as a fleet, which makes no sense.²¹³ The business venture being described here is a massive undertaking. The number of ships in this venture is more than the East India Company controlled in the eighteenth century, with considerable difficulties in ensuring that the ventures the company financed were used for the profit of the East India Company itself rather than for the personal benefit of the company's employees.²¹⁴ As Plutarch describes it Cato's interest is limited to lending, and a small equity participation, and the venture is being constructed largely to diversify his risk. Nevertheless the profits to him are described as large. Trading ventures can generate profits for the trader which are multiples of the amounts laid out, but interest rates on loans are at least an order of magnitude lower. How the profits of the venture were exactly distributed is not clear from the narrative. What is clear is that Cato formed this large company, exercised control over it and profited immensely.

To the extent that they enabled the separation of ownership and management of the business, and to the extent that their existence over a number of years encouraged reinvestment, these *societates* facilitated growth. Further groupings of individuals were family and other networks which spanned the Mediterranean and were involved in commerce, taxation and the movement of wealth.²¹⁵ Again, family groupings can be structures which extend over periods and also facilitate growth.

The presence of the elite can be felt throughout the trading business, as patrons and benefactors of shipping *collegia*, as lenders to traders and shippers and as organizers of commercial ventures in which they invested. Trade attracted opprobrium in at least some quarters, but could be the foundation of great fortunes. There were many ways in which lower ranking individuals could participate in this business, as small time shippers or traders or by building on commercial insights gained from service overseas. I believe that where the prerequisites of trade were in place the ease of entry into the business and the variety of entry points was such that trade flourished.

²¹³ That would have unnecessarily and dramatically limited the trading opportunities available and in any case a single representative could not adequately supervise or monitor the activities in such a fleet.

²¹⁴ It was difficult to monitor what the employees were up to when out of sight of head office. Keay (1993) 113 gives some figures for the East India Company which are relevant comparators. Between 1611 and 1620 the Company sent 55 ships to the east. Thirty years later the Company was sending less than half that number. The earlier ventures produced returns possibly as high as 155% per venture.

²¹⁵ See Rathbone (2003) 222.

4.7 Prerequisites

4.7.1 Information flow

Hopkins took the view that market information in the Roman world was limited and unreliable.²¹⁶ This view implies a more limited view on trade volumes. Robinson and Wilson take the view that most trade was directed rather than conducted on a tramping basis, and so implicitly assume that information flows were good.²¹⁷ The Piazzale delle Corporazioni at Ostia contained alcoves for corporations of merchants (in rope, leather, wood and grain) and also for shippers from different cities. I see structures such as this as being examples of how information flowed, and I see their construction as being at least partly motivated by the essential commercial interests of shippers and traders. Many ships left Rome empty and it was in the interest of shippers to fill their ships; it was in their interests to provide reliable information about demand in their cities to traders. It was in the interest of traders with goods to ship to find out where demand was best for their goods. In this area the trader had access to information from a wide geographic expanse and the shipper the opportunity to negotiate to fill his ship. The port was a clearing house for information. The Roman unification and pacification of the Mediterranean world greatly facilitated this flow and reduced previously existing barriers to trade.

4.7.2 State

A great variety of goods was traded but the state had only limited interests in much of what was traded. Its concerns lay principally with the supply of the army, the grain distribution at Rome, metals for coinage and the materials required for infrastructure. State demand played only a small part in the overall demand, and diminished proportionately as overall demand increased. For example, the state's interest in the provision of the *annona* at Rome, and elsewhere, and in the supply of the army amounted to only very small parts of total grain trade, very likely less than 5%.²¹⁸ In addition, state supply was very largely contracted out to private shippers.

²¹⁶ Garnsey (1983) 126-8.

²¹⁷ Robinson and Wilson (2011). Bang (2008) 136-40 also discusses some of the difficulties for traders in obtaining necessary information.

²¹⁸ If we take the number of recipients of the *annona* to be 300,000 of which 200,000 are at Rome and the balance elsewhere and assume an army size of 400,000 including auxiliaries and recognise that the urban population, which was perhaps 20% of a total population of 60 million, depended on trade in grain then the state interests are not much more than 5% of the total grain trade. The figure of 200,000 is based on Dio 77.1.1. I assume that most of the recipients of the *annona* are based at Rome and load this figure by 50% for illustration purposes. Significant but reasonable variations in that assumption will not lead to significant variations in the

The principal contribution of the state to trade was through the institutions which provided contract security, physical security, and currency stability and through the provision of physical infrastructure such as ports.

4.7.3 Money

The period of the first and second centuries AD was one where the currency was stable and there was negligible price inflation. Starting around the beginning of the third century AD the currency began to depreciate and towards the end of the third century hyperinflation had set in. The currency of imperial Rome was widely used throughout the territory under its control.²¹⁹ Monetary conditions for trade were ideal, particularly in the first two centuries AD.

4.7.4 Port structure

Sea transport requires significant infrastructural investment. The three main components of a port are firstly a harbour or shelter where ships can down anchor and rest, sometimes for extended periods.²²⁰ The technology to build harbours is ancient and the general design of harbours had been developed by the fifth century BC.²²¹ The second component is a wharf which is the land-sea interface across which cargo is transferred on and off ships. Transfer of cargo is either directly from incoming ships onto the wharf or from incoming ships onto smaller craft which discharge onto the wharf.²²² The third component are the silos where goods are stored. Silos are part of the process which manages the erratic flows of goods into the port and the demand for onward transmission.²²³ To the extent that the port is being used to receive goods for nearby consumption rather than for trading and onward transmission, it also needs a connection to the centre of consumption, whether by road or river canal. This is less of a requirement if the port is functioning more as a trading centre. The essential

conclusion that the state interests were not much more than 5% of the total. For example if there as many recipients of the *annona* outside Rome as within it then the total becomes under 7%. The figures for the army size are taken from Table 8.2.

²¹⁹ See Chapter 7.

²²⁰ At this part of its journey the cargo on board a ship is at its most valuable and protection is most desirable. Unprotected ports are less likely to be selected as destinations. But note that some ports can function without any harbour – for example some North African ports which consisted essentially of jetties jutting out into the seas far enough to reach beyond the shallows.

²²¹ Casson (1995) 365.

²²² Unloading was done either by beaching or by transfer onto lighters or by unloading directly from a ship onto a quay.

²²³ So, for example, at Ostia large parts of the storage capacity was used for the grain reserves of the city of Rome. Meiggs (1973) 276. It is also important to remember that the port is only part of the storage structure. The silos in the Egyptian villages and at Alexandria are part of the structure.

component of these structures in determining the volumes that can be processed is the wharf length.²²⁴ Growth rates depend on changes in aggregate wharf length. It is impossible to grow the volume of trade if wharf lengths do not increase. If the conditions for trade are favourable then an increase in wharf length will lead to increases in trade volumes. If conditions are unfavourable then trade will decline and port structures will no longer be maintained; wharf lengths will decrease.

The volume of goods that can flow through this structure in a day depends on the length of the wharf and the rate at which goods can be transferred per metre of wharf length per day (“transfer rate”). The transfer rate in turn depends on technology and work practices.²²⁵ It also depends on whether ships docked alongside the embankment or prow forwards, although it is generally easier to unload from the side. Since the transfer rate does not much vary through the period its quantification is not needed for an understanding of growth rates.

Quantification of the transfer rate is useful for understanding how much grain, for example, could be moved through a particular port.²²⁶ The following formula expresses the relationship between the amounts of goods that can be traded in a year, where at least part of the trading activity involves transport across the sea. Where the transfer from sea to land takes place through a harbour structure then the land sea interface is measured by the wharf length.

Where the transfer from sea to land takes place across a river quay then the land sea interface is measured by the length of the quay. Not that goods which transported by sea need to be

²²⁴ The question of how to measure port size is not straightforward. Measures include the size of the harbour basin, the length of the quays or jetties or wharf length. Stone (2014) 582 argues that the best measure of the size of a port is the length of the wharf. He argues this for two reasons. Firstly, it is a measure of the total length of all the different structures (jetties, quays etc.) across which goods are transported and so enables comparisons to be made between ports which have different structures. Secondly, wharf length is a measure of the total length of boats that can be docked while basin size is just a measure of the capacity of the port to provide shelter. I follow this approach because the essential function of the port structure is to provide an interface between land and sea and that interface is the wharf. I go further and argue that it is the best measure of economic activity in this economic structure, because of the limitations of the land based transport system.

²²⁵ Substantial changes in this variable only occurred in the 1960's. This happened with the advent of containerisation which dramatically reduced the time it takes to move goods on and off ships. A 20 tonne container can be moved in 2 hours. Containerisation reduces port loading and unloading times by about 90%. It reduces freight charges – which are largely related to loading and unloading – by a broadly comparable amount. It increases the processing capacity of the port eight times. This increase in processing capacity enabled the rise of China. Based on discussion with Phillip Skelton.

²²⁶ In terms of broad order of magnitude this figure is not difficult to estimate. Heinzlemann (2008) 5 indicates that a ship of 250 tonnes burden could be unloaded in between 6 to 8 days and Xx indicates an unloading rate of 30 tonnes per day, which are broadly consistent figures. Heinzlemann (2008) 5 indicates that along the Tiber a maximum of 20 ships were unloaded at any one time. Assuming a quay length of 1000 metres and 250 days of processing this implies a transfer rate of about 200 tonnes per metre, for ships berthed lengthwise. For ships berthed prow forwards I assume a transfer rate of 400 tonnes per metre. The essential question is how long it could take a team of men to unload a ship of so much tonnage manually (and in the case of certain cargo, for example stone columns, with the assistance of cranes) using technology that hardly varied until the middle of the 20th century.

processed twice, first from land to sea and then from sea to land. The volume of goods traded is therefore about half the total processing capacity of the system.

Processing Capacity = land sea interface x transfer rate per metre year Formula 4.2

I illustrate the effect of the change in infrastructure with reference to the principal ports and way stations of the system that supplied Rome and the surrounding area. Ostia, on the mouth of the Tiber, and Puteoli, in the Bay of Naples, were the principal ports of Rome through the Republic. By the time of Augustus Ostia had substantial unloading, loading and storage facilities, but no harbour.²²⁷ In the first century AD the harbour of Portus was constructed by Claudius and Nero and later much extended by Trajan.²²⁸ Through the first century and early second century there was a doubling of storage capacity at Ostia.²²⁹ Subsequently there was a continuing decline in the port with *horrea* and *tabernae* and *insulae* being abandoned at Ostia.²³⁰ Puteoli, which had been unaffected by the building of Portus, continued to function through the third century.²³¹ The following table illustrates the overall change in processing capacity that occurred.²³²

²²⁷ L. Casson (1965) 32.

²²⁸ Heinzleman (2008) 7 suggests that the structure of the storehouses at the new harbour of Portus indicate that it was designed for the daily needs of Rome while those at Ostia suggest that imports were, at least to an extent, received for trading purposes.

²²⁹ Keay (2008) 13.

²³⁰ Heinzleman (2008) 7.

²³¹ Keay (2008) 11.

²³² I have not been able to find estimates for wharf length for harbours near Rome similar in quality to those produced by Stone (2014) for North Africa and have therefore made rough estimates. Keay (2009) gives a harbour size of 200 hectares for the first harbour at Portus. The circular equivalent radius is 800 metres. Large parts of the circumference of the harbour are not used for unloading and loading. I have selected a wharf length of about half the circumference equivalent. Keay (2009) gives a harbour size of 32 hectares for the second harbour at Portus. The circular equivalent radius is 320 metres. Most of the circumference of this harbour is available for loading and unloading. I have selected a wharf length just under the approximate full length of the circumference.

Table 4.3: Change in processing capacity of principal ports around Rome

1:Port	2:Construction started under	3:Wharf Length (metres)
Puteoli	Republic	500
Ostia	Republic	1,000
Total end Republic		1,500
Portus First Harbour	Claudius	2,500
Antium	Nero	300
Portus Second Harbour	Trajan	2,000
Tarracina	Trajan	500
Centumcellae	Trajan	300
Total beginning of second century AD		7,100

The above provide estimates of changes in the **physical volumes** which could be processed through the ports around Rome. When the wharf length increased from 1,500 metres to 7,100 metres then the total physical tonnage that could be moved across the land sea interfaces increased proportionately. At the beginning of this period much of that capacity was absorbed in the importing of wheat, and the proportion of the capacity devoted to that commodity fell as capacity increased dramatically. To the extent that the new capacity was used to transport goods which were more valuable than wheat the increase in **monetary values** is more than the increase in **physical volumes**.

I discuss in Chapter 6 how considerations of status and of economic development led to infrastructure development and I will illustrate some of the interactions that led to such development. The basic processing capacity of the port system around Rome had increased about fivefold through the first and into the beginning of the second century. This represents a radical change in the economy of the city of Rome. I believe that part of the motivation for the construction of this new infrastructure was the difficulty in processing the city's basic needs through the infrastructure of the Republican system.²³³ Puteoli and Ostia were together large enough to receive the grain requirements of the city.²³⁴ The remaining processing capacity of these ports was barely sufficient for a third of the ships docking to be filled with materials for export. The provisioning of Rome and the other towns served by these ports was dependent on river supply, other small ports and beaches. As discussed in an earlier section, hinterland provisioning by land transport was sufficient only for towns of up to 10,000 inhabitants. Congestion risk was severe.²³⁵ The processing capacity of Puteoli and Ostia was so close to the basic needs of Rome that unevenness in the arrival patterns of ships into the ports must have caused bottlenecks and underutilization, and reduced the capacity of the ports to process basic requirements. Grain shortages might occur at Rome not because harvests had failed but because too many ships arrived at the same time. The building of the first harbour at Portus reduced that risk by almost trebling the capacity of the principal ports. The provision of this additional capacity had the ancillary effect of very substantially expanding the available trading opportunities since it reduced commercial trading risks; a ship carrying

²³³ Stone (2014) 595 attributes the building of the port structures in North Africa, which were immensely expensive, to local municipalities acting corporately and drawing on municipal revenues.

²³⁴ Taking Heinzlemann (2008) numbers of 20 ships berthed at any one time and assuming that it took between 6 and 8 days to unload a ship of 250 tonnes gives a total processing capacity at Ostia, if we assume 250 days operation per annum, of just over 200,000 tonnes per annum – this is the sum of exports and imports. The total processing capacity of Puteoli, which is smaller but ships might dock prow forward, was probably similar. The grain requirement for a million people was 300,000 tonnes.

²³⁵ A ship might spend between 3 and 4 weeks in its destination port. The length of time a ship spends in port was significant. A 250 tonne ship takes about 15 days to both unload and to load. Before getting to that processing stage it has to wait in the harbour, or outside the harbour, until it is allowed to dock – which also takes time. I use 250 tonnes as illustrative from a normal range of 60 tonnes to 400 tonnes because I am considering large towns. Because sailing times depend on the weather, and were therefore variable, the level of congestion in the system was difficult to manage. Indeed favourable conditions which brought a ship into harbour likely benefited other vessels and created congestion. A 250 tonne ship might therefore spend a total of 20 to 25 days in or near the port at which it discharged its cargo and reloaded. This is a substantial part of the journey time. Increases in this are a risk to shippers' and to traders' profits – the number of annual voyages might be reduced.

cargo bought for resale at Rome would find space to be unloaded. The growth in capacity continued and with it trade. The volumes traded were, in monetary terms, an order of magnitude higher at the end of the period than they had been at the beginning of the period.²³⁶ These values might have been 30 times higher at the end of the period. This amounts to an annual increase of over 3% per annum over a period of about 100 years, which is dramatic. The absence of any significant improvement in the port infrastructure around Rome in the second century indicates that trading volumes peaked there at the end of the first century

The story of substantial growth in port capacity indicated by Table 4.4 is mirrored in the other ports of the Empire as we would expect. The increase in importing capacity at Rome must be mirrored by an increase in exporting capacity elsewhere, and conversely.²³⁷ So, for example, we see a very marked and similar increase in port capacity along the African coast.²³⁸ This decline in trade does not appear to have happened along the African coast, when port construction appears to have continued into the third century but then to have effectively stopped.²³⁹ The continuing function of the North African ports after the decline in trading activity at Rome reflects the easy access from those ports to multiple markets. A decline in one part of the economy need not be catastrophic for other parts.

4.8 Conclusions

While the last two centuries BC saw a transformation in the economy through widespread monetization as will be discussed in Chapter 7, the first two centuries AD saw a transformative change in the level of trade. This change began from the middle of the first century AD when the infrastructure available for trade was dramatically extended. This extension may in part have been motivated by the state's need to facilitate the flow of basic goods. Extensive state expenditure, in excess of tax revenues, served to increase money supply and further facilitated trade. The engine for this growth was a social and legal structure which made available to men of any social standing the opportunity to create wealth for themselves through the purchase and sale of goods and their shipment, as price differences and differences in comparative advantage were exploited. There was a spreading of technology and knowledge which substantially changed trading patterns. The demands of

²³⁶ If we assume that the non-grain goods traded were higher value than the grain, say 5x higher, then the value of the goods traded at the end of the period is about 30 times the value of the goods traded at the beginning of the period.

²³⁷ Except to the extent that the port is being used for outward and return journeys of fishing vessels.

²³⁸ Stone (2014). The wharf length of Carthage is about 4,700 metres or about the same as the ports of Portus combined.

²³⁹ Stone (2014) 586.

the state and the elite had little effect on this substantial growth, although both benefitted from the transformation through increased tax revenues in one case and financial or commercial profits in the other. The increase in trade was associated with a general rise in the living standards of the non-elite. When the conditions for trade are good and there is at least a moderate amount of profit retained by traders, in other words when traders in the aggregate do not immediately consume all the profits from their ventures but reapply some of those profits into an increase in trading activity, then substantial growth in trade will occur over a reasonable period of time. During the first and second centuries AD there is a clear increase in the infrastructure available to support trading activities. I do not believe that these trading structures lay idle but rather that new harbour structures at Rome were used. From this it follows that in the period of the first and second centuries AD trade at Rome increased by about five times what it had been at the beginning of the period, since this is the increase in the processing capacity of the ports of Rome. This level of increase in trade over that period only required that traders reinvested relatively modest amounts of profits into the trading businesses.²⁴⁰ There is no reason to believe that the increased amounts of goods going into and out of cities such as Rome were purely related to the consumption of the elite. Such increased activity is likely to be associated with increased exploitation of comparative advantage, greater specialisation and improvements in the material culture of the population. The conditions for trade during the first and second centuries and into the third were ideal. Craftsmen who produced goods in the city of Rome, for example, for consumption there could as well sell those products to merchants for ultimate consumption at a great distance from Rome. Changes in trading infrastructure around Rome had implications for the productive centres located there; as trade became easier markets opened up for those who produced in the cities. In my view the extent of this productive capacity is an explanation for the survival of cities such as Alexandria, Antioch and Carthage following the withdrawal of their tax revenues and the consequent reduction in local demand supported by those revenues. Local demand was supplemented by overseas demand through the mechanism of trade. Most trade flowed into towns and cities across the wharf space of the Empire's ports and the length of wharf space is a powerful metric enabling quantification of changes in volumes traded by

²⁴⁰ If traders in that period each year reinvested 1% of turnover back into the trading business through increases in the amount of stock being transported then this fivefold increase would have been achieved in the period. What matters is that there is capacity to process this increase in stock across a land sea interface.

location and over time. This ability to track changes by location matters because changes in trade varied by location.

CHAPTER 5: The expectation of life as a measure of economic progress

5.1 Introduction

The life expectancy sustained within an economy is an extremely powerful measure of economic development and of social variations such as the level of inequality in the economy. The economic structure which generated agricultural surplus at a level which enabled large urban centres to produce goods which could be freely exchanged over a wide area raised the level of material culture. I will argue that the current dominant view that Roman life expectancy lay in the range 20 to 30 years and averaged about 25, is based not on an analysis of data specific to the Roman economy, but on the view that that was the range within which all pre-industrial life expectancy fell. As we have already seen economic conditions in pre-industrial societies varied. The economy of the Roman Empire was substantially more developed than that of medieval Europe. More nuanced approaches are possible and I outline one such approach in this chapter, which only requires limited data but significantly extends the insights that can be gained into economic performance. The clear indications are that economic performance was beyond the primitive state implied by current mortality estimates. Comparisons of life expectancy experienced during the time of imperial Rome with those experienced in Britain during the eighteenth century support the indications from the previous three chapters on the levels of agricultural production, of urbanisation and of trade.

5.2 The debate

Writing in 1966, Hopkins presented an estimate of Roman life expectancy of between 20 and 30 years.²⁴¹ The data available to him was over 4,000 inscriptions but his conclusions are not based on this data.

He argued against the usefulness of this data on two grounds. Firstly, because of what he considered were inadequacies in the data, and secondly because the mortality structures derived from the data did not fit to his *a-priori* expectations. The perceived inadequacies arose partly from his modelling approach. The subset of the data that he wanted to select was those inscriptions where the age of the deceased was recorded. So, if an inscription noted the relationship of the deceased to the person setting up the memorial but did not record the age of the deceased then he did not want to use that data. The data-set that he constructed had the

²⁴¹ Hopkins (1966).

defect that inscriptions set up by children to their parents were often not used because children tended not to include the age of their parents. His methodology needed age at death so for Hopkins the fact that length of marriage rather than age of spouse was often recorded was a defect in the data. Hopkins also noted (more generally) that infants were underrepresented while those aged 1 to 19 were overrepresented. He rejected the mortality table which he constructed from this data because it did not match any mortality table from a benchmark set of data, namely the Coale-Demeny tables which are described later in this chapter.²⁴²

Having rejected the available data as unusable, he formed a view on mortality based on general reasoning. Life expectancy must be at least as great as that needed to prevent the population from collapsing so it must be more than 20 years. Secondly, he believed that the features of the economy which determine mortality are wealth, the availability and quality of medical support, and whether the economy was agricultural or industrial. He assumed that an agricultural economy characterized by poverty and limited medical support would experience life expectancy of around 30 years. On this basis he concluded that life expectancy for the Roman economy was between 20 and 30 years.²⁴³

Writing in 1987 Hopkins further objected to the use of inscriptions giving the following reasons. Firstly the sample of data is biased because it relates only to those who could afford inscriptions or who were considered to deserve commemoration. As he recognized, this difficulty is not insuperable since the data can be read as representative of the class of individuals who were commemorated, so for example we might study the prosperous urban population. A more serious objection was that there were biases in the treatment of individuals within the class which commemorated its dead. Hopkins noted that young sons were commemorated more than young daughters, and wives more often than their husbands, and that infants under one year and young children were not sufficiently represented.²⁴⁴ It is, however, not sufficient argument against a dataset to note that there is underrepresentation. The underrepresentation may be statistically insignificant or it may be balanced by

²⁴² As illustrated in Section 5.4, a wide array of techniques has been now been developed for dealing with data imperfections and missing data and these influential conclusions of Hopkins should not be seen as meaning that the several thousands of surviving inscriptions have nothing to say about the quality of life of the individuals who were commemorated by them or about economic conditions generally.

²⁴³ Hopkins (1966).

²⁴⁴ Hopkins (1987).

overrepresentation, it may not affect the methodology used or it may be possible to adjust for the underrepresentation.²⁴⁵

Hopkins criticism is based on compilations of age at death data produced in the 1960's from many thousands of inscriptions.²⁴⁶ He notes patterns in these tabulations which make no sense and concludes that the surviving inscriptions cannot inform the debate on life expectancy. The conclusion as to the patterns which he observes in the tabulations are correct but his more general conclusion that inscriptions must be silent on the question of life expectancy is, however, not supported by this analysis. It is possible, for example, that the tabulations yield strange results because the process by which they have been produced is flawed.²⁴⁷ Attempts could also be made to understand and adjust for the weaknesses in the data, which is a standard approach. See later in this chapter.

In order to extract from the wealth of data insights into life expectancy I develop in this chapter an approach which can be used where age specificity in the data is unreliable. I have tested the theory which I have developed against different data sets and also considered its reasonableness from general reasoning. This approach is a significant development in demography generally. In particular it allows conclusions about how long people lived to be drawn from relationship data, as illustrated later. This powerful approach yields conclusions which make sense relative to the conclusions already drawn on agricultural performance, urbanisation and trade.

Hopkins formed the conclusion that mortality in the Roman world was between 20 and 30 years. This view was based not on any analysis of data but on an argument which extends over less than two paragraphs. The argument is 'On the assumption of a stationary population, the expectation of life at birth must have been above 20, because otherwise the difficulties of self-replacement are too great. Life expectancy at birth was also probably under 30, with infant mortality above 200 per thousand; for this has been generally true of pre-industrial populations and correlates with the predominance of agriculture, low average

²⁴⁵ See later in this chapter for description of datasets used to generate mortality tables and brief descriptions of how data inadequacies were dealt with.

²⁴⁶ The data that Hopkins uses are based on the work of Szilágyi (1961), Szilágyi (1962), Szilágyi (1963), Szilágyi (1965), Szilágyi (1966) and Szilágyi (1967) who constructed regional tabulations of numbers of male and female deaths from a review of many thousands of inscriptions.

²⁴⁷ These tabulations are flawed because, for example, the selection process limits the inscriptions included in the summarising process to those for which age at death is available. All inscriptions should have been included in the compilation process and at the very least a field of 'age unknown' should have been maintained. There may be other flaws.

income, and scarcity of doctors and of useful medical knowledge ...'²⁴⁸ In 1987 he observed that 'In general and in the long run, large pre-industrial populations have an average expectation of life at birth of between 20 and 40 years'.²⁴⁹ I believe that life expectancy in the Roman world was at the higher end of that range, and was somewhat beyond that for the elite section of the population.

In 1994 Bagnall and Frier attempted to construct mortality and population tables from Egyptian census data. The indications from this data were of mortality levels broadly consistent with that suggested by Hopkins.²⁵⁰ Frier in 2000 summarised the position as follows: 'Almost all historians now assume that Roman life expectancy at birth was approximately twenty-five years. This consensus rests ... on the reasonable assumption that, granted the general social and economic conditions prevailing in the Roman Empire, its life expectancy is likely to have been near the lowest levels attested for pre-modern populations.'²⁵¹ In 2001 and 2007 Scheidel argued that none of the available data could be used to inform the mortality debate and that the benchmark model life tables widely used were not appropriate. For example, he believed that the shortcomings in the Egyptian data severely limited their use (although it was possible to conclude that Egyptian females had a life expectancy of about 20 years), and also that skeletons could not be used to add to the debate on Roman demography because of the difficulties in estimating the ages at death from skeleton remains.²⁵²

5.3 The parameter P

Much use is made of benchmark data in the discussion of Roman demography and some of the debate centres on how applicable the benchmarks are. Because of the importance of these benchmarks this section discusses their construction and some of their limitations. In this section I also set out an alternative approach to the use of these benchmarks which substantially increases their explanatory power.

The commonly used benchmarks are the Coale and Demeny life tables.²⁵³ The published tables are the output of a formula which was generated from the mortality experience of a number of territories from the mid-nineteenth century onwards. The data represent a wide

²⁴⁸ Hopkins (1966) 261-2.

²⁴⁹ Hopkins (1987) 115.

²⁵⁰ Bagnall and Frier (1994).

²⁵¹ Frier (2000) 788.

²⁵² Scheidel (2001) and Scheidel (2007).

²⁵³ Coale and Demeny (1983).

range of mortality experiences and were grouped into four categories labelled North, South, East and West.²⁵⁴ The grouping for the Model West table is essentially Northern Europe and the principal English-speaking countries, with the addition of Taiwan, Japan and Israel. This grouping is the one most widely used by Roman demographers. The published tables for each grouping show standard life table values (such as probability of death at a particular age and life expectancies at that age) for 25 levels of mortality. The first level corresponds to a female life expectancy at birth of 20 years and the last level to a female life expectancy at birth of 80 years. These calculated tables are intended to represent a universe of possible mortality experiences.

It happens to be the case that this universe of possible mortality experiences can be described very simply by a single parameter as a consequence of the formula used. Recognition of this fact does not appear in the Roman demographic literature, and although there is some recognition of it in the demographic literature more generally, I do not think that this fact has been exploited previously to the extent that I propose.²⁵⁵ The parameter which I select to describe this universe is the probability, P , of surviving from birth to adulthood. I have generally taken the age at which the juvenile becomes an adult as 15 years but since mortality is broadly negligible in the teen years other ages, such as age 18, could be used with no effect on the conclusions drawn. The precise age used in the definition of transition to adulthood does not matter so long as it lies in a reasonable range.

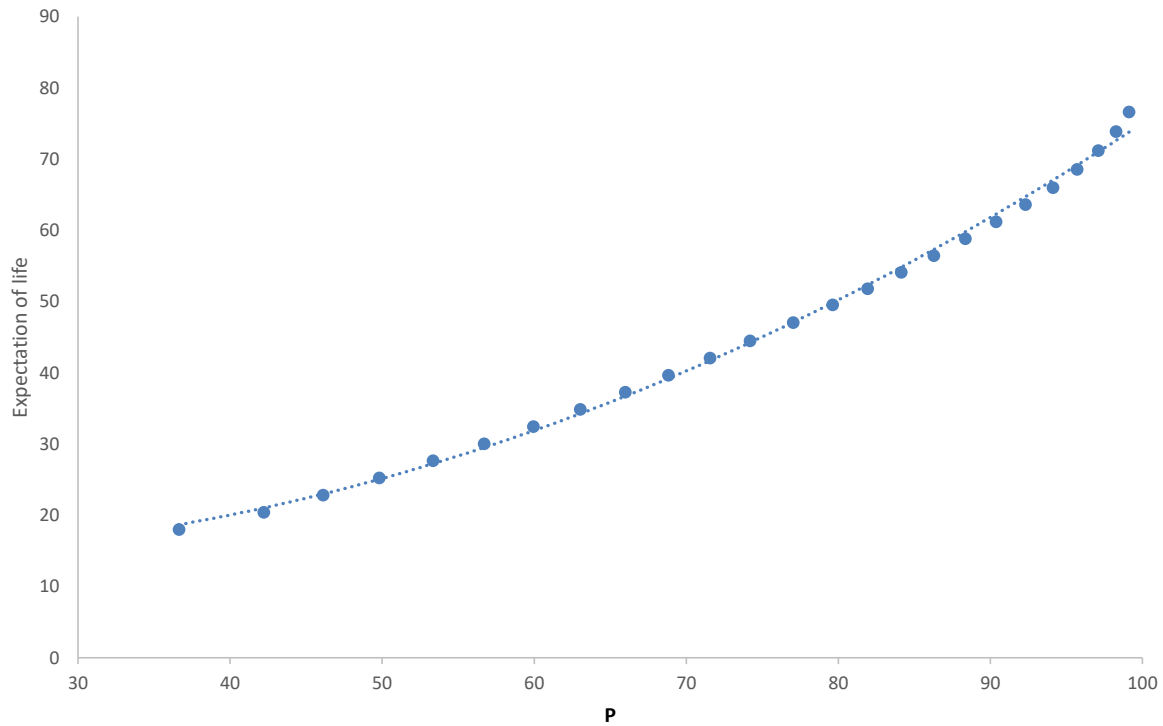
I select this parameter for three reasons. The first reason is that it represents the probability of a new born-offspring becoming capable of reproducing itself. A second reason for choosing this parameter is that it is an amalgam of infant mortality, childhood mortality and pre-adult mortality and therefore is likely to capture the effects of the mortality forces which operate across a wide age range. Finally, the level of data needed to estimate this parameter is substantially less than in approaches which attempt to estimate life expectancy from age specific exposure and death data. Indeed as will be demonstrated, no age-specific data need

²⁵⁴ The composition of the data which entered the model West category is as follows. Countries included are Australia, Belgium, Canada, Denmark, England and Wales, Estonia, Finland, France, Ireland, Israel, Japan, Latvia, Luxembourg, the Netherlands, New Zealand, Northern Ireland, Scotland, Sweden, Taiwan, South Africa and the United States. Note that not all of the life tables for each of these countries is included in this grouping as some of the Swedish mortality experience, for example, is included in the model North table. The table is described as being for those countries with 'good vital statistics' and is a 'residual collection'. The England and Wales tables are for 1871 to 1959.

²⁵⁵ In 1993 Woods reviewed the relationship between infant mortality and life expectancy at birth against a number of mortality data sets and found that generally they were so well correlated that one could be used to estimate the other. There were however cases, particularly where mortality was high, where the parameters could move independently. Woods (1993) 216.

be used. The following graph shows this probability P along the x-axis and life expectancy along the y-axis.

Figure 5.1: Relationship between P and the expectation of life: Model West

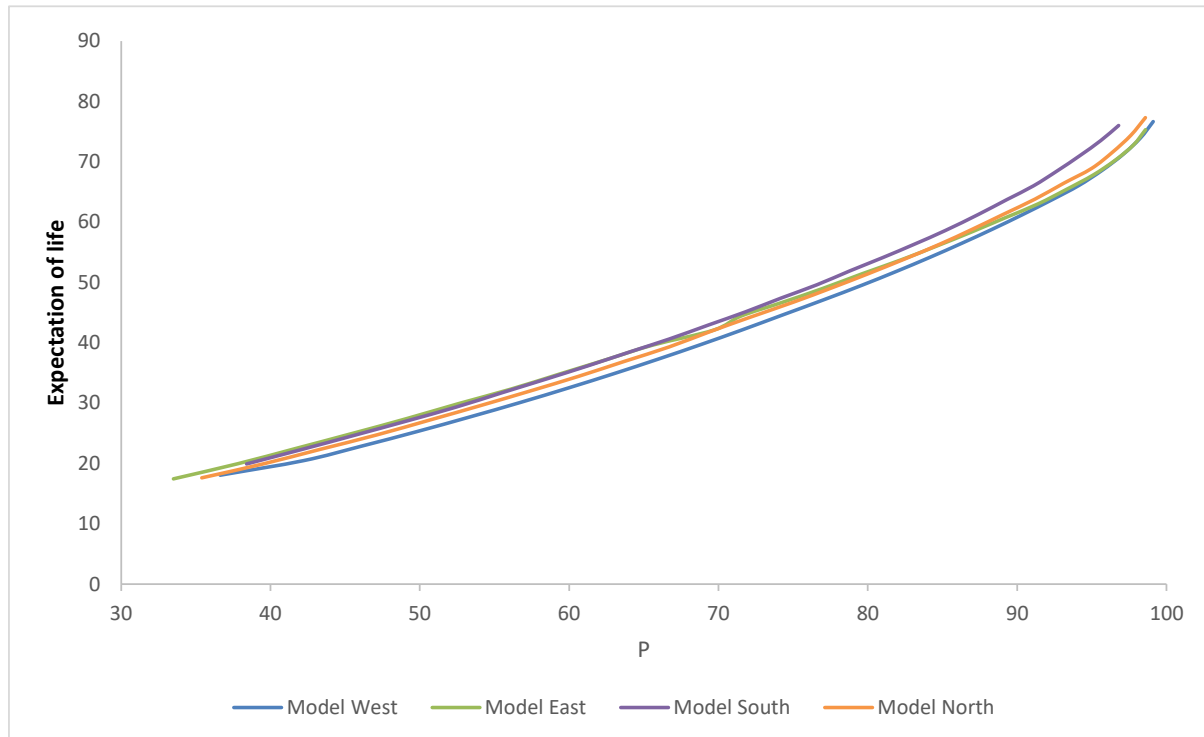


There is a one-to-one correspondence between the probability of surviving to be an adult and life expectancy. In other words if we know the probability of surviving to age 15 and we believe that the Model West life tables capture the range of possibilities then we know the life expectancy of the population. Indeed, because for each probability of surviving to age 15 there is only one Model West variant. Equally if we know that probability then we know the full age structure. In short it is not necessary to have detailed age-specific mortality data to construct a population structure.

This observation can be further generalized. The following graph overlays the four sets of life tables. To all intents and purposes the distinction between the groupings North, South, East and West is irrelevant for our purposes. The life expectancy corresponding to the same probability of surviving to age 15 is the same under the different groupings to within one or two years. This is an indistinguishable level of difference for the purposes of Roman demography where life expectancy can be scarcely determined to within a five year tolerance. To generate a population structure we need, therefore, only one statistic, provided that the

Coale-Demeny tables are appropriate and that we are interested in life expectancy to broad tolerances.

Figure 5.2 Relationship between P and expectation of life: All Models

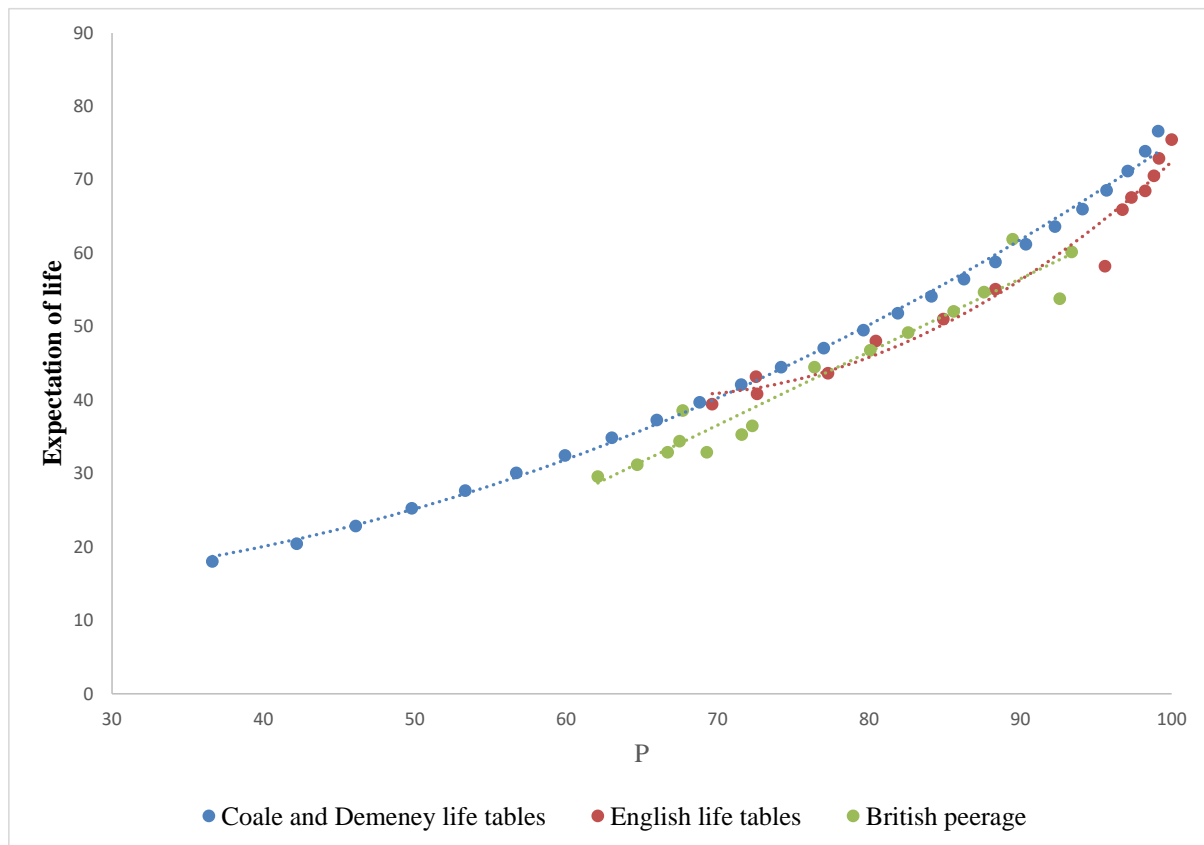


Writing in 2001 Scheidel argued against the use of the Coale Demeny benchmark data. He argued that the data was taken from low-mortality populations and that, because Roman demographers need to allow for high mortality regimes, they are using tables which are extrapolated from mortality experiences which differ to those in the economy they are investigating.²⁵⁶ Infectious disease, however, remained a significant cause of death even after the middle of the nineteenth century and is therefore reflected in the tables, and the very high level of mortality assumed for the low life expectancy in the Coale-Demeny tables necessarily implies a very high infectious disease component. I will also demonstrate that Roman mortality falls within the range of the benchmark data used to parameterise these tables. The tables are also used more generally by demographers as will be evident in the discussion of comparative data.

²⁵⁶ Scheidel (2001) 23-4.

The Coale-Demeny tables can be tested against more specific data. Here I choose two data sets one of which predates the data on which the Coale-Demeny tables are based. The data-sets are the mortality experience of the British peerage between 1550 and 1850 and the English life tables from 1850 to 1990. Note that the observed data for the British peerage includes a period of about 150 years when Britain continued to be afflicted by plague and by food shortages. In the following graph each possible combination of P and life expectancy under the Coale-Demeny table is shown. Each actual combination of P and life expectancy from the two data-sets is also shown.

Figure 5.3: Fit of British peerage data and English life tables to Coale and Demeny Model West



The observed data broadly matches that of the benchmark and indeed the fit is as good as we would expect any set of observations to have to a benchmark.²⁵⁷

²⁵⁷ The observations, however, almost all lie below the benchmark data which diverges from life expectancy in the British peerage in most years by up to 3 years, generally. In the case of the British peerage for the first three observations (for 1550 to 1600) the overstatement is 7 years and in the case of one observation at the end of the

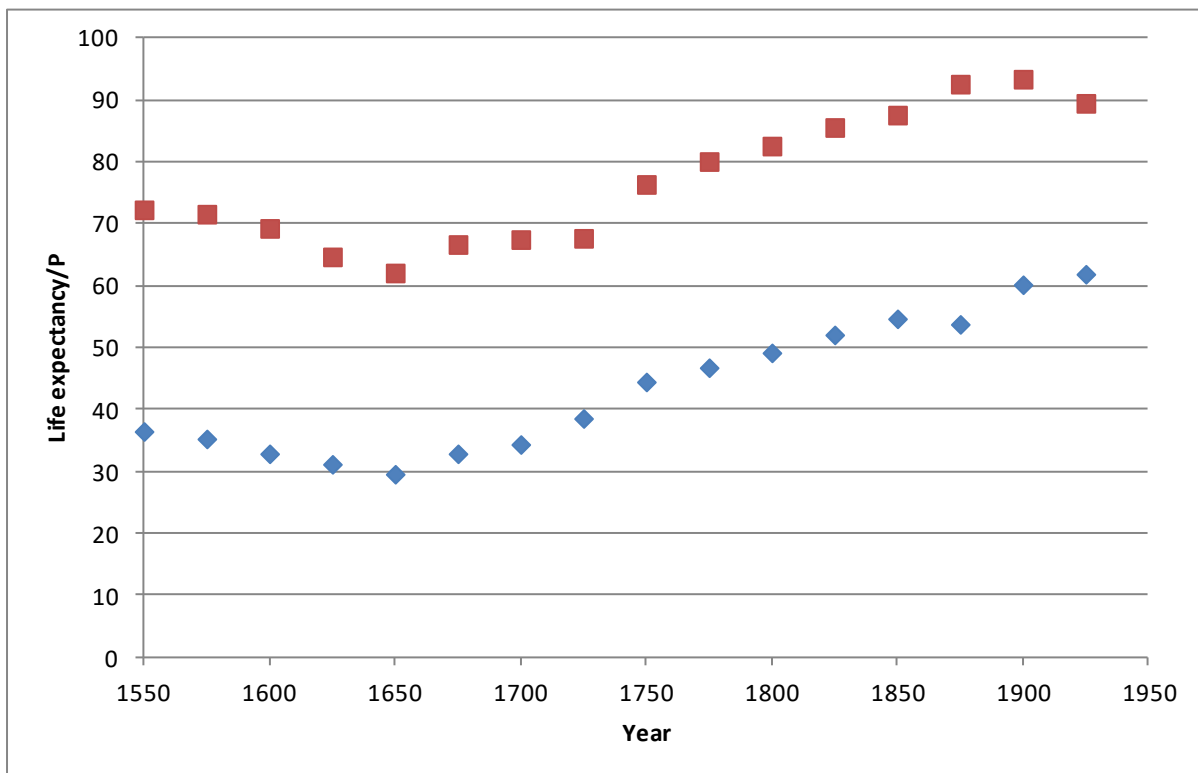
To reflect the level of accuracy that the benchmark is capable of giving, I quote all life expectancies derived from it to the nearest 5 years. Overall it is reasonable to conclude that the probability of becoming an adult is a powerful explanatory variable and also that the Coale-Demeny tables can be used to map that statistic onto a population structure to a broad level of accuracy.²⁵⁸

The following graph shows how the mortality experience of the British elite evolved over time and is used here to demonstrate further the relationship between P and life expectancy. Two sets of data are shown. The first set, which starts at 70% for the year 1550, is the probability that a newly born infant will survive to age 15, and the second set which starts at age 35 for the calendar year 1550 is the life expectancy. These have tracked each other over the almost 400-year observation period. By both measures mortality worsened from 1550 to 1650, and it was only in 1725 that the mortality experienced by the British peerage had returned to the level seen in 1550. The reasons for these changes are discussed in the next section of this chapter.

nineteenth century when early mortality fell substantially but life expectancy only marginally the difference is 12 years.

²⁵⁸ Using different datasets Woods found that the Coale-Demeny models fitted well to the data even when life expectancy at birth was less than 40. Woods (1993) 216.

Figure 5.4 Relationship between P and expectation of life for British peerage data



In short to determine life expectancy and population age structure we need only determine the probability of surviving to age 15.²⁵⁹ This probability can be determined, for a given population, in a number of ways. It is equal to

- the number of pre-adult deaths divided by the total number of deaths²⁶⁰
- the number alive at age 15 in any time period divided by the total number of births in that time period.

Note that in the first two of these cases it is not necessary to know the ages of the individuals. All that is necessary is that the individuals be classifiable into pre-adult and adult. Subject to the data set being not significantly biased then a complete age structure can be reconstructed from data which does not record age or which records age unreliably.

²⁵⁹ As already noted, we can estimate P by the probability of surviving to any teen year since the probability of death is typically at its lowest in the teen years.

²⁶⁰ In a stable population the number of deaths= the number of births and either quantity can be used.

5.4 Comparators and their implications

As has already been noted very limited progress has been made in deriving conclusions on mortality from available data. The common perception that life expectancy was about 25 years is based on a very limited review of comparator data and there is limited discussion on the need to distinguish between the different categories in the population or how their expectation of life can develop over time. This section attempts to look at comparator data in a way which makes the complexities clearer. It also attempts to show how almost all of the mortality datasets have imperfections but also that there is wide array of techniques to deal with these imperfections. I use a deliberately wide set of comparative data, from English parish records, registration set up for the taxation of Russian peasants, the records of the British peerage to English census data. Little of this data was originally compiled for demographic purposes.

The following table provides a range of comparative figures on life expectancy and mortality experience, using figures based on male lives. The sample is chosen to cover a very long time period and to illustrate the diversity of mortality experiences that are possible. The first observation is drawn from an analysis of Chinese rural data.²⁶¹ The period covered is 1929 to 1931, which was a period of civil war, when central government was weak and large sections of the country could not be reached. The data collected was census data and also event data; the event data counted the numbers of births, deaths and marriages in the previous year. The data had limitations. Age is not based on full years since date of birth but was counted as age 1 at birth and then increased by a year at each following lunar new year. Data related to deaths was substantially incomplete. Statistical techniques were used to adjust for these limitations. This analysis indicates extremely high mortality, relative to the Coale-Demeny tables, in the age range 5 to 20 which the authors attribute to a high level of tuberculosis. The second observation in the table again relates to China but is based on the genealogical records maintained by the Wang clan, which span the period from the first century AD to 1740.²⁶² Again the data contained limitations. There were a large number of incomplete records, dates of birth and death were not always available, and the data required substantial editing before it was used. Information related to juvenile deaths was almost completely absent – less than 1.5% compared to the authors' expectations of 30% to 40%. The authors estimated mortality at ages 30 and above from the data and considered a Coale-Demeny model provided a

²⁶¹ Barclay, Coale, Stato and Trussell (1976).

²⁶² Zhao (1997).

reasonable fit. On the assumption that the model applied at lower ages, a life expectancy of 34 years was indicated. The analysis indicated that mortality was broadly constant over the period analysed. The third and fourth observations in the table are based on the registers which capture information on all those subject to taxation in eighteenth and nineteenth century Moscow. These registers were periodically updated and record age of household member, date of entry into the household and date and cause of exit. The observations are based on the experience of serfs since most of the taxation was paid by peasants. Again the information was incomplete in some respects and it was supplemented by data taken from parish register where available for the villages being analysed. The information taken from the parish registers related to childhood deaths and to births to help in the estimation of infant mortality. Statistical techniques were used to adjust for missing data. Given the data difficulties the authors note that the results of the analysis agree reasonably well with the Coale-Demeny tables. The authors note that there was ‘stagnation’ in the development of mortality in the period up to the mid-nineteenth century and that this reflects the conditions of the serf population.²⁶³ The fifth through seventh observations relate to the mortality experience of France. These life expectancies were constructed from estimates of population distributions and from counts of numbers of deaths, with adjustment for age rounding in death declarations.²⁶⁴ The eighth through eleventh observations relate to the mortality experienced by the British peerage in the period 1550 through 1850. The data was divided into three universes. The primary universe related to British peers, the secondary universe to their offspring and the third to the next generation. The analysis was based largely on the secondary universe. Again there were imperfections in the data. Not all dates of death (particularly for the period up to 1750) were available and assumed ages at death were used.²⁶⁵ The last observation relates to the English population and is drawn from the English Life Tables.²⁶⁶ These are constructed following each new census and represent the mortality experienced in England and Wales. The census data is used to estimate the exposed to risk and the numbers of deaths between the most recent and the immediately prior census are taken from the national registers of deaths. Probabilities of death are then calculated.

²⁶³ Blum and Troitskaya (1997).

²⁶⁴ Blayo (1975).

²⁶⁵ Demography of the British Peerage (1964) 52-70. ???

²⁶⁶ Figures supplied by Adrian Gallop of the Government Actuary's Office.

Table 5.1 Illustration of the variety of mortality experiences

1: Population	2: Period	3: Life Expectancy at birth	4: Life Expectancy after one year	5: Infant Mortality Rate
1: Chinese farmers	1929-1931	24.1	na	na
2: Chinese Clan	1 st c AD - 1749	34.8	47.3	28%
3: Moscow serfs	1745 - 1763	24.3	35.5	34%
4: Moscow serfs	1851 - 1858	24.4	35.5	34%
5: French Population	1740 - 1749	23.8	33.6	32%
6: France Population	1750 - 1759	27.1	37.0	29%
7: France Population	1820 - 1829	38.3	46.6	20%
8: British Peerage	1550	38.0	na	na
9: British Peerage	1600	32.9	na	na
10: British Peerage	1700	34.4	na	na
11: British Peerage	1800	49.2	na	na
12: British Peerage	1850	54.7	na	na
13: English Population	1850	39.9	46.3	16%

The table makes clear some of the complexities of mortality regimes. For example, mortality can be a constant over time as the compilers of the data based on the Wang clan observed or it can show deterioration followed by continuing improvement as can be observed from the experience of the British Peerage.²⁶⁷ Mortality stagnation is evident for the Moscow serfs who over a period of a century show no improvement or deterioration in any of the key mortality statistics. This stagnation is not evident in the French population of the middle of

²⁶⁷ United States data also shows a pattern of mortality deterioration followed by mortality improvement. Pope (1992) 277 shows a decline in life expectancy in the United States in the late eighteenth century. The levels of life expectancy previously seen were not achieved until the end of the nineteenth century.

the eighteenth century. There is evidence of differences between elite and non-elite populations. The experience of rural China in the early part of the twentieth century is markedly worse than what the elite members of the Wang clan had long enjoyed. The generality of the English population in 1850 on average lived as long as members of the British peerage had in about 1750 and could expect to die 15 years earlier than their contemporary aristocratic neighbours. But their expectation of life was 15 years better than the rural poor of China in the beginning of the twentieth century.

These differences are to be expected. For example, an elite generally has access to a steady supply of food, it can secure warmth in winter and can be expected to have better sanitary conditions than the balance of the population. Equally it can choose where to live and to leave the city when it becomes unhealthy. We expect superior economic and social position to be associated with a better quality of life including having more of it.²⁶⁸

The table also shows the life expectancy after one year, where available. In all these cases where life expectancy at birth is 25 years the infant mortality rate is high. One third of infants die. On average those who survive the perils of the first year of life will be expected to live another 35 years and therefore to have about 20 years of adult life during which they can reproduce and labour. As we will see the reproductive capacity that this represents far exceeds what is needed for a population to sustain itself.

The progression of mortality for the British peerage shown in the last section serves as an example of the influences which shape mortality experience. Until about 1750 the British population was affected by mortality crisis years when there were abrupt increases in mortality from the normal underlying mortality. In the early modern period the most marked deviations from normal mortality occurred in the sixteenth and early seventeenth centuries, while the later mortality crises were less severe.²⁶⁹ Eight of the mortality crises of the period 1544 to 1666 were due to plague on its own or in combination with other diseases, and three, none plague related, can be associated with high food prices.²⁷⁰ Overall, however, mortality crises are generally not a major contributor to deaths and additional deaths were unlikely to have exceeded 10% of total deaths.²⁷¹ In the British population as a whole mortality

²⁶⁸ The correlation between economic status and wealth is recognised by insurance companies to a degree of subtlety. Heavier mortality assumptions were used for policies funded by premiums collected door to door than for policies paid for by direct debit.

²⁶⁹ Hinde (2003) 91.

²⁷⁰ Hinde (2003) 105.

²⁷¹ Hinde (2003) 97.

experience declined after the middle of the seventeenth century because of deterioration in both early childhood mortality and infant mortality.²⁷² Significant proportions of deaths in the period shown in Figure 5.4 relate to infectious diseases and it is the change in the level of these diseases that needs to be explained. The McKeown thesis attributes most of the reduction in the incidence of infectious diseases in this period to increases in the standards of living and nutrition.²⁷³ Although the theory has been criticized, and the causes for the improvement in mortality before 1850 are debated, no alternative explanation has been developed.²⁷⁴ I am not aware of any explanation for the reduction in the expectation of life in the period 1550 to 1650. This was a period of political instability and living standards and nutrition may have been affected as a consequence. In general I believe that there is a positive correlation between economic strength and life expectancy.²⁷⁵ Because of the high correlation between **P** and expectation of life I believe that **P** is also a measure of economic progress.

5.5 Relevant data and its interpretation

It has proven difficult to capture data that points as clearly at the underlying mortality experience of the Roman population as it has been for other populations already described. Some objections have been raised to data most notably by Hopkins but these objections are not insuperable. This section revisits some of the data sets to see what conclusions can be drawn from them and how they might compare to the comparator data described above.

5.5.1 Inscription data

I show below an analysis of inscription data which identifies how one might use data which does not include age information to generate a mortality table.²⁷⁶ In this table we count the numbers of times different categories of dedicator are shown on memorial inscriptions. The first line of the table shows the numbers of times husbands are shown as having set up the memorial. In these cases an adult death is being commemorated. In the third line I show the

²⁷² Hinde (2003) 102.

²⁷³ McKeown, Brown and Record (1982). See discussion Hinde (2003) 198-9. Frier (2000) 793 uses nutrition, sanitary conditions, compact urban living and the failure to adopt preventative measures such as quarantine as explanatory factors.

²⁷⁴ Hinde (2003). 199-203.

²⁷⁵ Changes in equality levels will also affect life expectancy levels. If there is an increase in economic value which is captured by a small part of the population the average improvement in life expectancy will be limited. Note also that the factors of production include human capital which incorporates knowledge of medical conditions and their treatments, techniques to manufacture drugs and the means to distribute this knowledge and care.

²⁷⁶ Using data taken from Edmondson (2005) 183-229.

numbers of cases in which parents set up the memorial. In these cases the death of a child is being commemorated. With the exception of the cases where the memorial was set up by parents I assume all cases relate to adult deaths. I calculate the probability of becoming an adult death as the number of adult deaths divided by the total number of deaths, with one adjustment.²⁷⁷ Because it is possible that there some underrepresentation of early life deaths in the data and to be conservative (i.e. understate life expectancy rather than to overstate it) I have assumed that 10% of the pre-adult deaths were not commemorated by a memorial. Note also that Hopkins indicated that inscriptions tended to underrepresent infants but to over represent those aged 1 to 19 and it is possible that these effects offset each other.²⁷⁸ Note also that very large numbers of children were commemorated and that with the exception of the figures given as for Republican Rome the relationship between the three quantities - number of husband and numbers of wives commemorating spouses and numbers of children commemorating parents – is reasonable. Because the Roman Republic data fails this reasonableness test I place very little weight on the conclusions drawn from that data. My assumption is that generally individuals who could afford to, and customarily did, commemorate their dead by setting up memorials did so for all family members.²⁷⁹

²⁷⁷ For a stable population numbers of births and numbers of deaths are equal. Adult deaths are deaths at age 15 or over. This is total deaths less deaths at age 15 or younger. This in turn equals total births less deaths at age 15 or younger. This equals numbers alive at age 15. Adult deaths divided by total number of deaths is then numbers alive at age 15 divided by numbers born. This is the probability of living to age 15.

²⁷⁸ Hopkins (1966).

²⁷⁹ See Graham and Carroll (2014), Gowland, Chamberlain and Redfern (2014) and Carroll (2014) for a discussion of infant commemoration.

Table 5.2 Illustration of a technique to estimate life expectancy with limited data

1: Relationship of dedicator to deceased	2: Rome - Republic	3: Rome non aristocracy	4: Noricum	5: Italy: Latium Regio XI	6: Lambaesis	7: Gallia Narborensis	8: Africa	9: Spain	10: Lusitania
Husband	54	24	33	30	24	23	14	15	15
Wife	11	15	19	10	12	9	14	13	13
Parents	15	31	25	24	23	23	35	34	34
Children	11	13	16	20	27	25	24	24	21
Siblings	7	10	3	9	9	9	8	9	8
Extended	2	6	6	7	4	6	5	5	8
Adult deaths	83	62	71	69	72	66	60	61	57
Pre Adult death	15	31	25	24	23	23	35	34	34
P	84%	66%	73%	74%	75%	74%	63%	64%	63%
Implied life expectancy	55	35	45	45	45	45	35	35	35

These figures for P (ignoring the results shown for Republican Rome) are very broadly consistent and indicate life expectancies of 35 to 45 years. The dataset relates to those who had the means routinely to set up inscriptions to their dead and we expect this subsection of the population to have had a favourable mortality experience. This expenditure on display is a discretionary outlay of wealth surplus to the need to maintain food supply and provide warmth and a clean living environment, and therefore indicates a higher expectation of life. Equally this expenditure can be seen as surplus to that needed to protect and care for offspring through the high risk periods of infancy and early childhood and to suggest lower mortality for those ages. The average life expectancy is about what was experienced by those born into the British peerage in the period 1700 to 1800.

5.5.2 Ulpian's evidence

The *Digest* preserves an opinion of Ulpian which are provided as benchmark valuation factors for annuities left in wills.²⁸⁰ The literature is confused in its understanding of these numbers. Both Hopkins and Scheidel, for example, describe these numbers as annuity values but then fail to incorporate the necessary consideration of interest rates into their arguments but rather treat them as undiscounted expectations of life.²⁸¹ This does not match the treatment of these numbers by those who used them and who were fully aware of the time-value of money.²⁸² Read in this light the Ulpian valuation factors provide strong evidence that mortality for the elite - who were those who were significantly involved in testamentary disputes - was light. The figures given by Ulpian for the valuation of an annuity to a man aged 28 is 30 times the annual income. This is the same valuation that is placed by Ulpian on a perpetuity. In modern annuity tables there is also little difference between the value of a life annuity for a 20 year old man and a perpetuity. I read this as indicating that the valuers of these annuities took the view that a man aged 20 would live very many more years, roughly between 40 and 50 further years. This is broadly equivalent to an expectation of life at birth of 60 years which is comparable to that achieved under the conditions enjoyed by the British peerage in the early nineteenth century. The figures attributed to Ulpian are intended to replace others which are described as customary figures. For age 60 Ulpian suggest a

²⁸⁰ These numbers are discussed further in Chapter 7 which deals with knowledge development in the Roman economy.

²⁸¹ Scheidel (2010a) 5 reads the Ulpian evidence to indicate life expectancy at birth of 20 years.

²⁸² See Chapter 6.

valuation figure of 5 against a valuation figure of 0 customarily used. It is possible that this reflects an adjustment to allow for improving mortality but if so it may not fully adjust for the improvement.²⁸³ In any case it is impossible to read these annuity factors against an assumption of high mortality. The implications are that the elite mortality was light – as the inscription data indicate.

5.5.3 Data from Roman Britain

The following table summarizes late Roman burial data and shows the stage in life at which the individuals died. The data is based on a sample of 30 late Roman sites which each contained more than twenty individuals.²⁸⁴ The percentages of deaths falling into the different stages of life groupings are very consistent across the three categories of urban, rural and nucleated and follow patterns which are consistent with *a priori* expectations. The authors of this paper present a mapping of these age categories to numerical age range; for example, they map young adults to the age range 18-25. I believe that this mapping substantially misstates the width of the age band. An interval of seven years for young adulthood implies an annual probability of death of between 2% and 4% which is too high and suggests that the age bands need to be widened.²⁸⁵ There are known difficulties with the aging of skeletons but these difficulties do not limit our ability to determine life expectancy from this data.²⁸⁶

²⁸³ The valuation factor used at age 20 would suggest greater expectation of life at age 60 than is implicit in the figure of 5 used by Ulpian.

²⁸⁴ Pitts and Griffin (2012).

²⁸⁵ Unless the populations were largely military which is not the case.

²⁸⁶ Contra Scheidel (2010) 13 ‘However, unless skeletons can be made to inform us reliably about mean age at death they will never reveal life expectancy per se.’

Table 5.3 Example of the use of cemetery data to derive life expectancy without age data**Percentage of deaths in different age bands**

1: Age Category	2: Urban	3: Nucleated	4: Rural
Juvenile	31%	27%	29%
Young adult	16%	21%	15%
Middle adult	38%	34%	34%
Old adult	15%	18%	21%
P	69%	73%	71%
Implied Life Expectancy	40	45	40

The proportion of juvenile deaths is about 30% which implies a probability of becoming an adult of about 70% or a life expectancy at birth of about 40 years. This data-set is also consistent with the comparator data and my interpretation of the inscription data discussed above. It may, however, be drawn from a wider social spectrum than the memorial data and provides an indicator that in this territory at least life expectancy at birth was generally lower therefore than for the sections of the population who erected memorials.

5.5.4 The Egyptian data

The following table is based on a compilation of the data from surviving Egyptian census returns. The surviving census returns imply large amounts of missing data and what is presented in the table below does not incorporate any attempt to complete the missing data. The table shows numbers of lives in different age bands.²⁸⁷

²⁸⁷ Bagnall and Frier (1994) 75-109.

Table 5.4 Difficulty of fitting Egyptian census data to a population structure**Percentage of the census population in different age bands**

Age range	Male Village	Male Metropolis		Female Village	Female Metropolis
0-4	24	26		28	8
10-14	7	16		24	14
15-19	13	23		21	11
Total	169	181		211	126

There are some difficulties with this dataset. For a population which is stable, or changing at only a moderate rate the numbers at successive ages should decrease monotonically. This feature is not replicated in three out of four of the samples. Equally we expect to see a male: female ratio of about 1:1. This feature is also not replicated. The differences between the data and *a priori* expectations are almost certainly statistically significant and strongly suggest that either the original census collection exercise or the process by which the census data have survived was biased. Nevertheless the Egyptian census data have been treated as sufficiently credible to enable life expectancies to be derived. These have been stated at 20 to 25 years for the female population and at more than 20 years for the male population. Because of the difficulties with the data, these conclusions cannot stand on their own and do not weaken the conclusions drawn from comparator data.²⁸⁸

The following data is a summary of numbers of deaths in successive age cohorts as recorded either in mummy labels or on inscriptions.²⁸⁹

²⁸⁸ There are further difficulties with this data. Taken at face value this data set indicates birth rates of about 3% which for a stable population implies an average age at death of over 30 and a life expectancy of about 50 years at birth.

²⁸⁹ Scheidel (2001a) 31.

Table 5.5 Further difficulty of fitting Egyptian data to a population structure**Percentage of deaths in different age bands**

Age range	Upper Egypt Mummy Labels	Upper Egypt Inscriptions	Lower Egypt Inscriptions
0-19	38.5%	24.6%	35.6%
20-29	23.5%	19.3%	19.0%
30-49	19.9%	24.7%	24.4%
49+	18.1%	31.3%	20.8%

This data is problematic. There are too many deaths in the age range 20 to 29 compared to the age range 30 and over. If this data-set provides unbiased samples of deaths over age 20 then it implies an average annual mortality rate for those aged from 20 and 29 of between 3% and 5%.²⁹⁰ This is about 300% to 500% of that implied by the Coale-Demeny life table with the most extreme mortality and is in any case an unbelievable characterization of the mortality levels for individuals in the prime of life and from a class sufficiently wealthy to afford expensive memorials. It is also close to the level of mortality that is experienced nowadays by people in their mid-to late seventies.

5.6 The application of P to population growth rates

In order to understand population dynamics we need only understand three parameters.²⁹¹ The derivation of P and its broad quantification has already been dealt with. The second parameter GRR, which is the gross reproduction rate or the average number of births over her lifetime to a woman who has attained the age of 15. The relationship between these two parameters and G, the intergenerational growth rate is as follows. If P is the probability that a newly born child will survive to age 15 and since about half the births are female each female aged 15 replaces herself with on average this number of females aged 15

²⁹⁰ In a stable population the total number alive at age 20 equals the total number of deaths at ages over 20. The probability of surviving from age 20 to age 30 is the total number of deaths above 30 divided by total number of deaths above 20. The average annual probability of survival from 20 to 30 is the 10th root of that quantity. The complement of that is the average annual probability of dying in the age range 20 to 30.

²⁹¹ Ignoring the effects of migration.

P.GRR.(0.5)**Formula 5.1**

and the population growth rate over a generation is then

P.GRR.(0.5)-1**Formula 5.2**

The following table takes P at 70% and then shows the interrelationship between GRR and intergenerational growth. At this level of mortality the difference between a GRR of just over 2 and a GRR of just over 4 is the difference between population collapse and substantial growth.²⁹² There is some relationship between these figures and those included in the Augustan marriage laws which exempted men and women from certain duties if they had at least three children.²⁹³

Table 5.6 Relationship between births and intergenerational growth rates

1: Gross Reproduction Rate (no of births)	2: Intergenerational Growth Rate (change in population size)
2.0	-30%
2.5	-25%
3.0	5%
3.5	23%
4.0	40%

Note that there is a social need to keep population size up. A reduction in births will in time reduce the availability of men for military service and will also reduce the availability of men for productive work. Both of these have clear negative implications for the security of the group as a whole and specifically for those who control the group and have continuing interest in their personal survival. In such cases they will tend to act to correct the population decline. GRR is an average of the age specific fertility rate from age 15 to age 50.²⁹⁴ Age specific fertility rates decline with age and increases in the age at which women marry pushes reproduction further into the lower levels of fertility and can dramatically reduce the GRR.

²⁹² As a consequence it is generally not difficult for populations to recover from substantial loss – provided there is adequate food.

²⁹³ Higher thresholds were used elsewhere.

²⁹⁴ The weights are the product of the probability of being alive at each age and the probability of being married – assuming that illegitimate births are ignorable. For a large part of this age range mortality is level and GRR is not very sensitive to variations in that factor.

Encouraging earlier marriage and placing social stigma on those females who fail to marry can move the reproductive population back into the range of high fertility. These changes can have dramatic effects on fertility. For example, a 25% reduction in the number of women getting married reduces fertility by 50%.²⁹⁵ This variation in GRR is enough to move from the very high population growth shown in Table 5.7 to population collapse. Moderate changes, therefore, in the rate at which people get married have major effects on population growth rate and can be managed by marriage laws or other customs which encourage early or late marriage. Individuals or social custom have the potential to further manage population size. For example, continued breast feeding and therefore lactation postpones ovulation and acts as a contraceptive.

5.7 Conclusions

I take the Coale-Demeny tables to be valuable benchmarks even for plague-afflicted populations and despite the criticisms levelled at them.

I have introduced a powerful methodology which needs limited data to draw conclusions on the life expectancy of a population. Rather than attempt to get detailed age specific tabulations of population density or deaths it is sufficient to get broad information on one parameter. The parameter I use is P , the probability of becoming an adult. This determines life expectancy for all the Coale-Demeny tables. In other words, you do not need age distribution data to determine life expectancy. All that is needed is roughly the proportion of childhood deaths or the rough proportion of the population who are in their teens and from these we then get life expectancy. Generally speaking the probability of surviving to age 15 is 70% in the Empire. This parameter can also be used to link different types of evidence. So if, for example, literary evidence gives us information on GRR and archaeological evidence gives us information on population growth then we can get to life expectancy via the parameter P , because of the interrelationship between the three parameters. P is also a measure of economic progress and the value of attempting to estimate it lies partly in the insight it gives us into economic progress. Skeleton data, therefore, can be highly informative as to economic progress.

There are difficulties with the data available for the estimation of P or of life expectancy but that is no different to the situation elsewhere in demography and more attempts could be

²⁹⁵ Hinde (2003) 132.

made to rectify the data limitations. The data produced from the Egyptian sources, however, have severe limitations. The population was not stressed but showed low voluntary growth rates through the Republic followed by higher growth rates during the Empire.

The fact that we need only estimate **P**, the probability that a new born child becomes an adult, in order to know life expectancy to a reasonable degree of accuracy means that data which cannot be used directly to construct life tables can be used to generate conclusions on life expectancy. The available data indicate that the life expectancy of the elite of imperial Rome was comparable to that of the elite in England in the eighteenth century, supporting conclusions drawn from a consideration of agricultural productivity, urbanisation and trade.

CHAPTER 6: Productive knowledge, the unrealised potential for growth and the social institutions which sustained the economy.

6.1 Introduction

The purpose of this chapter is to incorporate discussion of the knowledge base of Imperial Rome into discussion of its economy, to enhance comparison to other economies and to identify factors which limited change. As we have already seen the total level of economic output depends on the relative productivity of labour engaged in non-agricultural activities, which in turn depends on the level of skills. The form of knowledge that I am concerned with is that which contributes to the production of goods and services. Such knowledge includes knowledge that makes labour more productive and also includes knowledge that is embedded in infrastructure such as ports or capital goods such as ships. The skills involved in valuing financial assets or surveying sites for engineering projects are examples of relevant productive knowledge. This knowledge facilitated the economic development outlined in the previous chapters.

I am not concerned, in this chapter, with knowledge which does not contribute to economic activity. So for example, facts enumerated in Pliny's *Naturalis Historia* which are interesting or fascinating, but do not contribute to economic activity, are not part of the knowledge base that is considered here. I will, however, consider the extent to which economic development can be furthered by an experimental approach driven by curiosity.

The forms of knowledge that we are concerned with include those acquired through schooling but also what is acquired through apprenticeship and more informally. Task-specific knowledge was important to the development of the economy and furthered by personal contact between individuals. Literacy and numeracy skills which enable symbols to be recognised and used in a commercial context but which are insufficient for the enjoyment of literary texts are economically important.

6.2 The debate

Writing in 2012 Saller attempted to locate Roman human capital in the spectrum that had preceded it and that succeeded it. He followed the view that much of the development in twentieth century economy came from increases in the level of human capital rather than to changes in the other factors of production. He categorised the routes by which that capital was developed as being through training received through family, training received through

apprenticeship and training received formally in schools or the equivalent. Almost all of the rural training he put in the first category, as a broad assumption. He broadly assumed that formal education was exclusively urban and limited to the artisan class and the elite. He principally used his view on the proportion of the population which received formal education to benchmark the Roman economy and concluded that Roman investment in education was high relative to pre-modern investment and low relative to early modern developed economies. He noted the role that cities play in the dissemination of knowledge because of the contact between individuals in more densely populated centres.²⁹⁶

Earlier work by Woolf, in 1999, emphasised the broader ways in which forms of literacy and numeracy contribute to the economy and that literacy and numeracy were pervasive. For example, the process by which grain produced in Egypt was consumed in Rome was a series of stages from production on an estate through river transportation to Alexandria and from there to the harbours of Rome and eventual distribution in the city. Each of these stages involved the production and storage of receipts, the use of documentary controls and many of the stages involved the production of accounts.²⁹⁷ Similarly, olive oil amphorae from Spain were marked and stamped to indicate weight when empty and full and where the contents came from. Skilled personnel at the harbour side knew where to look for and how to interpret the markings and stamps.²⁹⁸ What had been created was ‘a complex urban world, one where the essentials of life had to be obtained through the retail market, where many sold their labour, rented accommodation, and were as a consequence functionally numerate. Public notions of time, weights and measures, and the calendar, ordered daily life, as law and custom ordered the city.’²⁹⁹

The level of formal education other than schooling is illustrated by Bradley in 1991 who gives a survey of thirty Roman Egyptian apprenticeship documents which cover the period from the end of the first century BC to the third century AD. Almost all of these contracts relate to weaving or other activities involved in cloth making but there some single instances of contracts relating to other occupations such as nail making, flute playing, copper smithing,

²⁹⁶ Saller (2012).

²⁹⁷ Woolf (1999) 62. Contrast Finley (1999) 181 ‘Graeco Roman bookkeeping was exceedingly primitive, restricted to a listing of receipts and expenditures ...’

²⁹⁸ Woolf (1999) 57, 59.

²⁹⁹ Woolf (1999) 51.

and shorthand. The apprenticeships are generally for between two and five years duration and it is likely that apprentices started their training around the age of twelve or thirteen.³⁰⁰

6.3 Knowledge, labour and growth

Classically growth is linked to investment spending, to education and research and development and the level of human capital is classically measured by years of schooling. Although schooling may have been the primary way in which knowledge contributed to economic growth in developed countries in the twentieth century, measuring levels of knowledge by years of schooling understates the importance that knowledge makes to the production of goods and services.³⁰¹ Formal schooling may play a limited role and informal apprenticeships may be far more significant.

Growth may also occur because of increases in productivity which result simply from the dispersal of knowledge. The mere contact between two individuals one of whom has skills that the other lacks and which can be profitably employed can lead to an increase in the overall knowledge base and to an increase in the production of goods and services. Dispersal of knowledge may also be as important as innovation. In other words it is not just the generation of new knowledge that matters but the moving out of that knowledge or old knowledge through the economy. This dispersal may happen informally through contact between individuals in urban centres or in structures such as the army.

Knowledge is not only conducive to economic growth by changing the contribution that labour can make but may lead to higher output through infrastructure, in other words through the generation of capital rather than the transformation of labour. As already noted the steam engine led to a transformation in the size of the land based hinterland of towns and it therefore increased the maximum possible size of the towns.³⁰² Higher levels of urbanisation also lead, provided that urban labour is more productive than rural labour, to higher levels of total economic output.³⁰³

³⁰⁰ Bradley (1991) 107-8. It is fascinating that almost all of these contracts are to do with cloth making. It is possible that this activity, which leaves extremely limited archaeological trace, was a significant component of economic output, as it is in India, for example, today. Horden and Purcell (2000) 354-9 discuss textiles and note that about a third of the Price Edict sections relate to textiles and that the items are 'highly differentiated'.

³⁰¹ See Saller (2008).

³⁰² The construction of the Great Western Railway brought Bristol within the daily range of London. Note there is evidence that rail tracks were used at Ostia and engines based on steam power may have been used to draw wagons, although there is no evidence that the use of either technology was widespread. Morley (2000).

³⁰³ On reasonable assumptions as to what that level of urbanisation might be we can say that the knowledge embedded in the steam engine enabled a substantial increase in economic output. For example if, the steam engine increased urbanisation from 30% to 70% then it effectively quadrupled economic output, regardless of

Knowledge that is developed out of simple curiosity may become economically relevant although it can take some times for such knowledge to become economically significant. I illustrate this by reference to the development of electromagnetism. From around the beginning of the eighteenth century a process of knowledge development took place, largely in Europe. This process resulted in the accumulation of knowledge as outcomes from experiments were shared among individuals working separately in scattered locations.³⁰⁴ Ideas as to the possible relationships between these different observations were formulated and also shared. These relationships were often expressed in mathematical language which had been developed in England and Bavaria in the seventeenth century and theories were developed which provided structures for explaining the observed phenomena. A vocabulary emerged that assisted in the expression of new ideas and helped in the assimilation of relationships between concepts. This vocabulary, often build on metaphors or by reference to the names of individuals who had made significant contribution to the body of knowledge, includes such words as ‘current’, ‘field’, ‘volt’ and ‘ampere’.

Basic facts about electricity and magnetism had been known since antiquity as, for example, the fact that amber when rubbed with wool acquires the ability to attract objects. Towards the end of the eighteenth century Coulomb, in France, formulated a rule which describes the attraction and repulsion of two bodies statically charged in a way similar to amber. At the beginning of the nineteenth century Volta, in Italy, produced the first battery and generated a flow of electric charge. By 1820 Ørsted, in Denmark, had noticed that the flow of electric current past a compass needle made the needle move and thereby discovered a connection between the two previously disparate phenomena of electricity and magnetism. Ampere, in France, added to this by discovering that two parallel wires carrying electric current attract or repel each other depending on the direction in which the current was flowing and behave like magnets. Shortly after this Faraday, in England, constructed an apparatus in which a wire rotated around a magnet when electric current flowed through the wire. He later found that moving a magnet though a loop of wire could cause a current to flow through that wire. Mechanical motion could generate electric current and electric current could generate

the level of relative productivity. If we take relative urban productivity at 4.0 then it moves relative output from 2.7 to 10.3. Or 1.6 to 6 if we use a factor of 1.5 See Appendix A.2. Also see Chapter 2 for illustration of the importance of the land based transport system daily range.

³⁰⁴ The social backgrounds of the individuals who participated in the development of knowledge in this period were also diverse. The grandfathers of Henry Cavendish, who discovered hydrogen, were the Duke of Kent and the Duke of Devonshire. Michael Faraday, who made significant experimental contributions to electromagnetism, was apprenticed to a bookbinder.

mechanical motion. Since electric current could be sent along wires it was now possible to build systems in which mechanical motion at one point could be used to generate mechanical motion at another point with almost no limit on the distance between these points. For example, water flowing over a dam could now be used to turn wheels many miles distant from the location of the dam. There was nothing in the material culture of imperial Rome to prevent similar developments. For whatever reason this particular process of experimentation, theorising, language development and of sharing never emerged during the time of the Roman Empire and the economic developments that it enabled, as unintended consequences, also never emerged.

When the knowledge accumulated though this process first entered the economic sphere its contribution grew exponentially as successful applications were reproduced and new applications were developed. As, for example, it was discovered that electric current going through a filament could be used to generate light then increasing amounts of work were expended on the creation of light bulbs, more were produced and more demanded. With the increased demand for electric current more work was spent in the production of current. Separately it was discovered that current could be used to generate sound and a similar process then happened with the telephone. The process of economic development was two-way exponential. There was exponential growth in the number of applications and there was exponential growth in the use of each of those applications. This viral growth was supported by a language through which the basic underlying knowledge could be communicated, including the language of current, charge and electrical resistance, a belief in the reality of the phenomena being described and an environment which caused individuals to work to produce ever more practical applications and manufacture ever more artefacts. Although the concepts used in theories of electromagnetism and the devices constructed by experimenters never emerged in imperial Rome, I believe that this exponential process of development and specialisation occurred in the area of trade, as already discussed.

The next three subsections consider three examples of the types of knowledge which were available in the time of imperial Rome. The examples are drawn from the fields of engineering, finance and agriculture and are used to illustrate how knowledge was used and how it failed to develop. The examples are also used to illustrate the features of the economy which meant that knowledge led to economic development.

6.4 Financial knowledge

Productive knowledge can reside as much in language as in the apprehension of facts and the common understanding and use of a specific technical vocabulary can increase the productive capability of labour. It can also exist in the common acceptance of certain relationships such as between the value of one unit of money now and a unit of money later. In the area of modern finance the technical vocabulary includes words such as ‘interest’, ‘present value’, ‘discounting’ the meaning and use of which is acquired through formal training and through apprenticeships. This language enables things to be done or to influence what takes place. It is applied to the valuation of assets ranging from life annuities to properties to government bonds and stock market companies and is used to describe techniques such as discounted cashflow. The existence of this language and these techniques support financial markets which contribute to modern economies. Much of this capability existed within the economy of Imperial Rome as evidenced by the Digest. The Digest makes clear that in Imperial Rome it was a working assumption that a monetary value could be put on almost any asset, whether physical or not.

I will use a small number of passages to illustrate these points. The first passage makes clear that valuations of future streams of income are to allow for the timing of the payments and for interest. It also makes clear that the present value of a future payment is less than the future payment.

*circa legem Falcidiam in eo, quod sub condicione vel in diem alicui relictum est, hoc observandum est: si decem sub condicione alicui fuerint relictæ eaque condicio post decennium forte extiterit, non videntur decem huic legata, sed minus decem, quia intervallum temporis et interusurium huius spatii facit quantitatem decem.*³⁰⁵

There is this to be noted concerning the *lex Falcidia* where something is left under a condition or for a given period; should ten be left to someone under a condition which is realised after, say, ten years, it is not a sum of ten which is regarded as having been bequeathed but something less; for the time lapse and interest of the interval reduce the sum from ten.³⁰⁶

This passage deals with the *Lex Falcidia* which was enacted in 40 B.C. In essence this law said that heirs must get at least a quarter of the estate to which they were heir and it required

³⁰⁵ Digest 35.2.66. See also Digest 15.1.9.8.

³⁰⁶ Using the translation of Watson.

for its operation that all the assets of the estate be valued. All assets of the estates were subject to this valuation regardless of whether they were tangible, such as a tract of land, or intangible, such as a right of way.³⁰⁷ If the heir to an estate thought that he had not been left the share due to him under the Falcidian law then an arbiter was appointed to determine the values of the assets and liabilities; the heir notified the other beneficiaries who presented their cases.³⁰⁸ Arguments presented to the arbiter need to reflect the time value of money.

The second passage provides a norm for the interest rate to be used in valuations for testamentary purposes.

item si rei publicae in annos singulos legatum sit, cum de lege Falcidia quaeratur, Marcellus putat tantum videri legatum, quantum sufficiat sorti ad usuras trientes eius summae, quae legata est, colligendas.

Likewise, if a legacy payable annually to a municipality is bequeathed, and a question arises with reference to the Falcidian Law, Marcellus thinks that only as much should be considered to have been bequeathed as will amount to a sum which, at four per cent interest, will provide the annual payments of the legacy.³⁰⁹

The valuation approach used here is exactly that used in valuing real estate based on rental income and a norm for rental yield. It also tells you how to do the valuation. It says that calculations can provide the answer. This is an important theoretical position.

The third passage provides another valuation basis and this passage when read together with the previous passage forces the realisation that there is a connection between interest rates and market values.

*cum Titio in annos singulos dena legata sunt et iudex legis Falcidia rationem inter heredem et alios legatarios habeat, vivo quidem Titio tanti litem aestimare debeat, quantum venire id legatum potest, in incerto posito, quamdiu victurus sit Titius.*³¹⁰

If the legacy of ten (whatever) a year was left to Titius and the judge had to work out the lex Falcidia as between the heir and the legatees, while Titius is still alive he must

³⁰⁷ See Digest 35.2.23 for reference to valuation of a right of way for the purposes of the Falcidian law.

³⁰⁸ Digest 35.3.1 (6).

³⁰⁹ Digest 35.2.3 (2), translation from Scott (1932).

³¹⁰ Digest 35.2.55.

reckon it as what the legacy would sell for, it not being known how long Titius will live.³¹¹

Market values rather than a formula can be used for assessing the value of income streams. This passage implies the existence of a market in these financial assets which was sufficiently deep as to be useful for valuation purposes.³¹² There is documentary evidence that *alimenta* could be bought and sold.³¹³ In AD 180 Turbo, an athlete who had won the right to two maintenances at the public expense in boxing matches which were held at Antinoupolis sold those rights to a local councillor who bought them for the benefit of his two sons, minors at that time. The sale was effected through a bank which transferred HS 4,000 from the councillor to the illiterate athlete on whose behalf Horion, a weaver, signed the deed of sale.³¹⁴ The sale of estates, encumbered by *alimenta*, also provided a source of market values. Comparing the value of such an estate with the value of a similar but unencumbered estate gives a value of the *alimenta*. The valuation approach used here is the same as the market value approach used nowadays in valuing companies for merger and acquisition purposes. A fourth, complex, passage presents a way of thinking of the valuation of ownership and the valuation of use:

si tibi decem milia legata fuerint, mihi eorundem decem milium usus fructus, fient quidem tua tota decem milia: sed mihi quinque numerari debebunt ita, ut tibi caueam tempore mortis meae aut capitis deminutionis restitutum iri. nam et si fundus tibi legatus fuisset et mihi eiusdem fundi usus fructus, haberes tu quidem totius fundi proprietatem, sed partem cum usu fructu, partem sine usu fructu, et non heredi, sed tibi cauere boni uiri arbitratu. Sed si duobus eorundem decem milium usus fructus legatus fuerit, quina milia accipient et inuicem et heredi satisdabunt.

If ten thousand *aurei* are bequeathed to you and the usufruct of the same ten thousand to me, the entire ten thousand will belong to you; but five thousand must be paid to me on condition that I give security to you that, ‘at the time of my death or loss of civil rights, they will be delivered to you’. For if a tract of land is devised to you, and the usufruct of the same land to me, you would indeed, have the ownership of the

³¹¹ Translation kindly supplied to me by Professor John Crook in 2004.

³¹² Digest 7.1.12 (2) says that the usufructuary can sell his rights, giving further evidence of a market in these rights. Digest 35.2.23 uses *pretium viae* as the value of a right of way for the purposes of the Falcidian.

³¹³ P.Lond. III 165 1164(i), translated with useful comments in Johnson (1959) 397-8.

³¹⁴ There are analogies between this market and nineteenth century markets in reversionary interests in estates.

entire tract, but you would have part of it together with the usufruct and part of it without and I should give security which would be approved by a good citizen to you and not to the heir. But where the usufruct of the same ten thousand *aurei* is bequeathed to two persons, they will each receive five thousand, and must give security to one another and to the heir.³¹⁵

Finally a fifth passage gives a series of factors which illustrate the valuation capabilities that existed and the subtlety to the approaches used, which were practical.³¹⁶ It addresses the problem of how to place an economic value on a stream of income which will be paid to an individual only as long as they are alive.³¹⁷ I interpret these valuation factors as legal norms which were used for capitalising future streams of income and which allow for life expectancy and for the time value of money. The passage says that where, for example, an individual aged twenty five has been left a yearly sum equivalent to HS 100, say, which is to be paid to him for the rest of his life then that stream of income is to be valued at HS 3,000.³¹⁸ The same income left to a man aged fifty five, who has fewer remaining years of life, is to be valued at HS 700. The value of a perpetuity of HS 100 is set at thirty times one year's income or HS 3,000. Income streams from a perpetuity and a young life continue sufficiently far into the future as to merit, for practical purposes, the same financial value.

libro secundo ad legem vicesimam hereditatium. computationi in alimentis faciendae hanc formam esse Ulpianus scribit, ut a prima aetate usque ad annum vicesimum quantitas alimentorum triginta annorum computetur eiusque quantitatis Falcidia praestetur, ab annis vero viginti usque ad annum vicesimum quintum annorum viginti

³¹⁵ Digest 7.5.6.

³¹⁶ The subtlety and sophistication of the thinking is illustrated by the following aspects. The passage explicitly refers to the valuation of a maintenance and a usufruct and makes clear that the same valuation approach is being used for these two instruments correctly implying that arguments could be made that different approaches could be used. The valuation rate of interest is $3\frac{1}{3}\%$ which is less than the 4% used earlier for the valuation of a perpetuity. A usufruct or a maintenance is a real asset with income expressed in real rather than monetary terms and a lower rate of interest is appropriate.

³¹⁷ These numbers have long been incorrectly read as reporting a compilation of life expectancies in Imperial Rome telling us how long the average Roman citizen was expected to live, and although this interpretation creates difficulties, it is the consensus view. For the use of the *Digest* passage as a summary of life expectancies see Roby (1886) 188-91, Burn (1953) 2-31, Duncan-Jones (1990) 93-104, Frier (1982) 213-51, Saller (1994), Parkin (1992) and T. De Vries & W. Zwalve (2003) 277-98. Trenerry (1926) is a significant exception and treats the Ulpian numbers as capital values of annuities. Greenwood (1940) 246 reads the tables as legal norms for valuations. Stein, (1962) 335-56, limiting himself to the maximum figure of 30 shown in the text, takes this figure to be used because it is the time between one generation and the next. I believe that to interpret these as valuation factors which allow for life contingency but do not allow for the other factors, particularly interest income, which were considered in valuations is incorrect.

³¹⁸ Since these are valuation factors rather than life expectancies the implied remaining life expectancy of a twenty five year old is much more than thirty years.

octo, ab annis viginti quinque usque ad annos triginta annorum viginti quinque, ab annis triginta usque ad annos triginta quinque annorum viginti duo, ab annis triginta quinque usque ad annos quadraginta usque ad annos quadraginta annorum viginti, ab annis quadraginta usque ad annos quinquaginta tot annorum computation fit, quot aetati euis ad annum sexagesimum deerit remisso uno anno: ab anno vero quinquagesimo usque ad annum quinquagesimum quintum annorum novem, ab annis quinquaginta quinque usque ad annum sexagesimum annorum septem, ab annis sexaginta, cuiuscumque aetatis sit, annorum quinque. eoque nos iure uti Ulpianus ait et circa computationem usus fructus faciendam. solitum est tamen a prima aetate usque ad annum trigesimum computationem annorum triginta fieri, ab annis vero triginta tot annorum computationem inire, quot ad annum sexagesimum deesse videntur. numquam ergo amplius quam triginta annorum computatio initur. sic denique et si rei publicae usus fructus legetur, sive simpliciter sive ad ludos, triginta annorum computatio fit.³¹⁹

The following is Watson's translation with my corrections.

The Five Percent Succession Tax Regulations Book 2: For computations to be made in the matter of maintenance, Ulpian gives the formula: From birth up to but not including the twentieth year, the amount of thirty years maintenance will be assessed and the Falcidian quota thereof be due; from twenty up to but not including twenty-five, the amount is twenty-eight; from twenty-five up to but not including thirty, the amount of twenty-five; from thirty up to but not including thirty-five, the amount of twenty-two; from thirty-five up to but not including forty, the amount of twenty; from forty up to but not including fifty let there be a computation of as many years as are lacking to sixty, with one year's remission; from fifty up to but not including fifty-five, the amount of nine years; from fifty-five up to but not including sixty, the amount of seven years; from age sixty no matter what the age, five years. Ulpian says that we observe this rule also for the computation of a usufruct. Still, it has been the practice for the computation from birth to thirty to be thirty years but from thirty of as many years as are lacking of sixty. The computation never goes beyond thirty. So equally in the case of a legacy of a usufruct to the state, whether for the provision of games or without restriction, the valuation will be of thirty years.

³¹⁹ Digest 35.2.68

The title of this work by Aemilius Macer refers to the five percent inheritance tax established by the *lex Julia de vicesima hereditatium* which was enacted by Augustus in AD 6 to contribute to the funding of army discharge bounties and created an inheritance tax of initially 5%, doubled by Caracalla and restored to 5% by Macrinus, which was imposed on estates of Roman citizens. Its application had been amended over time and it came to be applied only to estates over a certain size and did not apply to legacies left to close relatives. The inheritance tax law required a valuation of the estate assets so that the state's share of one twentieth could be determined and the formula that is set out in this passage was also used in the calculation of estate taxes.³²⁰

There was in some sense no need for these valuation factors. The courts decided in each particular case the value of the various assets of the estate by reference to what might actually or theoretically be achieved in the market for similar assets, adjusting for any anomalies in market conditions. The value of these benchmarks might flow from the fact that substantial amounts of work might be needed for the valuation of complex assets which were minor parts of the estate. There was also the difficulty that a testator wanting to satisfy obligations to many friends and relatives and constructing a will with a large number of legacies might need to know how much these might be valued at for the purposes of the Falcidian law and the inheritance tax. I believe that the numbers being presented in Digest 35.2.68 are broad summaries of what actually happened when evaluations were carried out in practice and were prepared as benchmarks to help judges and others evaluate estates. They are summaries of past practice to be used as precedents and are presented as a formula. These numbers could not have been more than benchmarks. They might also have been used to identify market anomalies which might lead to unfair estate valuations. Prices can fluctuate over time and the same article might fetch different prices in different places; the *Digest* indicates that in such cases the approach used was an adjusted market value approach with anomalous values removed, and these benchmarks may have helped in the process.³²¹

How the valuation factors might have been constructed is a matter of debate. One answer to this question is that the numbers were compiled through a statistical analysis of ages of death

³²⁰ For a discussion of this tax see Gilliam (1952), 397-405 and Neesen (1980) 135-40. Frier (1982) 223 argues that two different sets of numbers were used in estate valuations. Frier (1982) 216 considers that for Falcidian purposes a capitalized annuity is intended but that the numbers in the text are life expectancies. This requires him to assume that 'the new use that the compilers of the *Digest* put Ulpian's life table was an improper use'. This then allows him to use these figures as life expectancies. I believe this approach is mistaken.

³²¹ Digest 35.2.63 (2) refers to differences in prices for the same commodity at different times and places. The literary sources also show an awareness of the concept of a fair price. Pliny *Ep* 1.24, for example.

recorded in cemeteries or by funeral clubs.³²² It has, however, been objected that no evidence for such a Roman statistical capability exists.³²³ There was, in fact, no need to compile detailed statistical data to produce these numbers, it was sufficient to observe trading patterns in the assets to which these numbers relate, or court decisions in valuation disputes. This is no different to modern practices. The valuation multiples that we use as benchmarks for property valuations are collations from numbers of examples. We observe the relationship between the prices of different properties and the rents achieved on those or similar properties and form a view, that for a particular class of property, the relationship normally works out at around a certain number and we take that number as our benchmark. Or we observe the relationship between market values and square footage and proceed similarly. We then use these benchmarks to estimate what other properties might sell or rent for. Similar procedures are used in the valuation of stock market companies in merger and acquisition transactions. There is very little to distinguish the valuation approaches and philosophy used as a practical instrument and set out in the Digest from that which underlies modern financial transactions in real estate, bonds and stocks.³²⁴

6.5 Engineering knowledge

In this subsection I illustrate the extent to which engineering knowledge contributed to the economy and in particular to consider how the structures that existed allowed this knowledge to be used. The engineering skills involved in the building of infrastructure such as canals and aqueducts were deployed through administrative structures which could identify local needs. In effect specialist skills which were rarely needed in any one location could be deployed across a wide area. These administrative structures were peopled by individuals who were motivated to develop infrastructure which in turn enhanced trade and economic development. The successful development of infrastructure was associated with enhanced status.

6.5.1 Tacit knowledge

Much of the knowledge needed was tacit. Even in modern economies much knowledge cannot be codified and can only be acquired through personal contact. For example, significant parts of the knowledge needed to test and build nuclear weapons is tacit and is acquired by personal contact rather than from documents. This knowledge is preserved by the

³²² Stein (1962) 44 mentions funeral club data. Frier (1982) assumes that some data related to deaths was used.

³²³ Hopkins (1987) sets out arguments against the assumption that graveyard data was used. Parkin (1992) 35-41 also discusses data difficulties.

³²⁴ I exclude derivatives and I specifically draw comparison with transactions.

doing of things and within local centres where that personal contact happens. Although the principles of how to construct fission and fusion bombs is public in practice the building of nuclear weapons is an art and it 'rests on knowledge that has not been, and perhaps could not be, codified'. Much of the difficulties come from the practical engineering problems, such as how to make crucibles which do not dissolve when molten metal is poured into them. The craft of the practitioner rather than just his knowledge of atomic theories is what matters. Moreover, there are endless possibilities for what might happen in different circumstances and only a limited number of experiments that can be carried out. Such possibilities include what happens when the weapon ages, when the surrounding temperatures are somewhat different, if it is manufactured under different processes. As important for the reproduction of a nuclear capability as the theory that underlies the design of nuclear weapons is the mere fact that it is generally known that such weapons can be constructed, transported to where they are intended to be used and will explode when needed and not before. Much of the process of reproducing a nuclear capability is one of reinvention and trial and error and that process is more likely to be successful if the project sponsor and workers know that it can be done.³²⁵ This applies more generally in professions.³²⁶

Writing in 2001 Collins further illustrates the importance of tacit knowledge and gives an explanation in the context of experimental science and the difficulty of reproducing experimental results.³²⁷ Scientists in Moscow University were the first to produce a measurement of the Q of sapphire. Their results could not, however, be reproduced and were not believed. The eventual reproduction of the experiment followed on from personal contact between the original experimenters in Moscow and those seeking to reproduce the experiment in Edinburgh and on the building up of trust between the parties. Written descriptions of the experiments were not sufficient. Collins illustrates the difficulties of knowledge transmission which does not involve personal contact between individuals and trust between the individuals. There are an extremely large number of things that the novice

³²⁵ Mackenzie and Spinardi (1995).

³²⁶ Formal education is only a part of the process needed to develop human capital. Informal training, effectively an apprenticeship, remains a significant part of modern professional training. For example, the completion of the formal exams of the Institute and Faculty of Actuaries is not considered sufficient for members to undertake work of high responsibility, such as the certification of Lloyds syndicates as being adequately funded. Evidence of extensive practical experience in working in the field, effectively under the supervision of senior experienced practitioners, is required before members are permitted to provide certification.

³²⁷ The specific context is the measurement of the quality of Sapphire which first took place in Russia and was then only reproduced in the West twenty years later. Striking a material, such as the metal of a bell, causes it to vibrate and these vibrations die down after time. The length of time that it takes these vibrations to die down to half their original strength is called the Q of that material. How to measure this factor is well known in theory.

could concentrate on in attempting to successfully carry out a new task. Absent personal contact between the expert and the novice the expert does not know which of these are being concentrated on and the novice does not know which questions to ask. The expert may not have fully formulated a theory as to which aspects mattered and why and indeed it may not be possible to formulate a theory which is of any use in passing the skill from the expert to the novice.³²⁸ Collins also emphasises the importance of trust in leading to reproducibility of results and illustrates the point by reference to the trust that existed in the social class structure in England.³²⁹

7.3.2 Infrastructure development

In this subsection I illustrate the deployment of knowledge through the Roman Empire in the construction of infrastructure which supported trade and urbanisation. The following passage is part of the dialogue between the Pliny as governor of a province and the emperor Trajan.

*est in Nicomedensiom finibus amplissimus lacus. per hunc marmora fructus ligna materiae et sumptu modico et labore usque ad viam navibus, inde magno labore maiore impendio vehiculis ad mare devehuntur ...hoc opus multas manus poscit. at eae porro non desunt. nam et in agris magna copia est hominum et maxima in civitate, certaue spes omnes libentissime adgressuros opu omnibus fructuosom. superest ut tu libratores vel architectum si tibi videbiter mittas, qui diligenter exploret, sitne lacus altior mari, quem artifices regionis huius quadraginta cubitis altiore esse contendunt. ego per eadem loca invenio fossam a rege percussam, sed incertum utrum ad colligendum umorem circumiacentium agrorum an ad committendum flumini lacum; est enim imperfect. hoc quique dubium, intercepto rege mortalitate an desperato operis effectum. sed hoc ipso (feres enim me ambitiosum pro tua gloria) incitor et accendor, ut cupiam peragi a te quam tantum coeperunt reges.*³³⁰

There is a sizeable lake not far from Nicomedia, across which marble, farm produce, wood and timber for building are easily and cheaply brought by boat as far as the main road; after which everything has to be taken on to the sea by cart, with great difficulty and increased expenditure. (To connect the lake with the sea) would require

³²⁸ So, for example, in learning to ride a bicycle or in learning to play the piano or learning to speak.

³²⁹ Collins (2001).

³³⁰ Pliny, *Ep* 10.41. See also Pliny, *Ep* 10.42, 10.61 and 10.62.

a great deal of labour but there is no lack of it. There are plenty of people in the countryside, and many more in the town, and it seems certain that they will all gladly help with a scheme which will benefit them all. It remains for you to send an engineer or an architect, if you think fit, to make an accurate survey and determine whether the lake is above sea-level. The local experts say that it is forty cubits above. I have looked at the site myself and find there is a canal dug by one of the former kings of Bithynia, though whether this was intended to drain the surrounding fields or to connect the river I am not sure; it was left unfinished, again I cannot say if this was because the king died suddenly or despaired of finishing the work. This, however, only fires me with enthusiasm to see you accomplish things kings could only attempt; you will forgive my ambition for your greater glory.³³¹

This passage illustrates the operation of the political and administrative system, during the very early part of the second century AD, in the production of infrastructure and, while there is interest in status considerations, the project is directed to economic ends. The members of the status group of which Trajan and Pliny are both members may not involve themselves directly in trade but status considerations encourage them to develop infrastructure necessary to trade. The project is intended to improve connectivity and to make local trade easier. Reducing the cost of transport of goods and making the process easier will lead to an increase in the volumes of marble, agricultural product and other commodities that can be traded. The factor limiting this improvement in infrastructure is knowledge rather than labour which Pliny indicates is plentiful and what is needed for this local project to start is an architect or engineer with the necessary survey skills. The passage illustrates how dialogue between the provincial governor and the emperor was used to mobilise the necessary expertise and how the administrative structure allows expert knowledge to be deployed across a wide area. This structure allows localities to benefit from knowledge generated at a great distance from themselves. The interests of the local population in generating trading opportunities for themselves are furthered by the centrally appointed local governor in dialogue with the central administration. This structure, which shares information on local needs and which deploys necessary expertise, facilitates the construction of extensive infrastructure for very limited burden on the revenues collected from taxation. The continuing deployment of architects, for example, to different projects also increases the expertise of these men as well as spreading knowledge to local areas. This level of knowledge enhancement is difficult

³³¹ Loeb translation.

when the connections between the centre and the local areas is disrupted and such disruption is harmful to trade.

This process of deployment of necessary knowledge is further illustrated by the construction of a water supply system at Saldae in about the middle of the second century AD aspects of which are described on a monument set up by the engineer responsible for the work, which celebrates his pride in his achievement. Saldae was to be supplied by water flowing first through a tunnel in a mountain and then flowing above ground into the town. The plans for the route of the aqueduct were drawn up by Nonius Datus, a legionary seconded to the project who very likely acquired his engineering skills in the army. Nonius Datus was later recalled because of problems which had occurred in the construction of the tunnel. The work involved two teams of men digging from each end of the tunnel to meet in the middle but both tunnel segments had been dug veering to right and therefore did not meet. Nonius was given management of the project and completed the work. The memorial which Nonius Datus set up to celebrate his achievement, and which includes the words of a letter detailing the request that he be put in charge of the project, has on its surviving sides the goddesses *Patientia*, *Virtus* and *Spes* which reflect the capacities to continue on with a task even when things are difficult, the ability to do the work and the hope that it will be successful. He describes how he motivated the men working on the project by setting up a competition between the navy men and the mercenaries.³³²

As with the construction of the canal the success of this infrastructure project depended on the appropriate technical knowledge, and project management skills, being brought to bear on it. This technical knowledge was not locally resident and would not be expected to be since these projects are not carried out frequently in one location but are carried out frequently across a wide area. Economic development was furthered during the time of imperial Rome through the deployment of knowledge which was supported by administrative structures which operated over a wide geographic area and by a status system which valued infrastructure development.

6.5.1 Agricultural knowledge

While large parts of the developments that have occurred in modern times were beyond the reach of the Roman economy, not all were. Further developments in agriculture, for example,

³³² Cuomo (2011).

lay within reach. Yields on the durum wheat grown in Roman times were limited because this type of wheat has a long stem and weak straw. The figure on the next page illustrates the difference in wheat types. This characteristic leads to lodging, the stem collapses and the grain lies on the ground, if the grain weight increases significantly, thereby causing loss of yield and quality, or even making the crop impossible to harvest. These genetic characteristics set a ceiling on the yield that can be obtained. In fact, an attempt to increase yield through improved fertilisation, for example, after some point increases the risk of lodging and so is counterproductive.

The improvement required to increase yields to modern levels was to first change the genetic characteristics of the plant so that it was capable of carrying the higher yields without collapse. This was achieved by genetically improving straw strength, shortening it and improving disease resistance.³³³ Further genetic developments to enable increased absorption of natural and artificial fertilizer were then of potential benefit.³³⁴ The development of shorter stemmed wheat types was the first step in the process. The production of shorter stemmed wheat types does, however, have the disadvantage in ancient economies that it reduces the quantity of straw available and this material was used in bedding and building in our period. Nevertheless, the start in the process beyond what was achieved in Roman times was the improvement in seed selection.

³³³ The factors contributing to the improvement are numerous. Substantial improvements have occurred in the last century in land productivity because of 'improved varieties, mechanization, more effective pest and disease control, better production patterns and improved land management'. Dixon et al (2006) 489. Nevertheless the principal enabling factor is the shortening of the stalk.

³³⁴ About half the size of modern yields is attributable to chemical fertilizers. Stewart et al (2005) 1. The improvement in grain types since the nineteenth century has increased grain yields by about 60%. Austin et al (1989) 295. There is an interaction between these two factors: modern grain types are efficient when placed in fertile soils under modern land management but prove less effective when placed on poorly managed and less fertile land. Modern developments also include chemical pesticides and herbicides. These latter developments have the advantage of reducing the amount of labour required directly on land management.

Figure 6.1: Modern wheat more productive than ancient wheat because the stalk is shorter



Parts of this selection process were available in Roman times including through observation of differences in wheat types in different parts of the Empire. The widespread use of one wheat type may have limited awareness of different performance characteristics of different species and the possibility of improving performance by breeding. Methods for seed selection were, in any case, also only taken so far.³³⁵ The approach used was to simply select the best grain from the crop and use it as seed for the next year.³³⁶ While there is evidence in the agronomists of some attempt to improve the wheat stock, for example by importing grains from overseas, there is also some evidence that this attempt at improvement was resisted, as in the case of the passage by Pliny cited earlier.³³⁷ The failure to understand the importance of short stemmed wheat is also evident in Pliny's reference to the millet imported from India into Italy which could grow to seven feet.³³⁸ There was, therefore, no general appreciation of the fact that a scientific approach to breeding and seed selection could improve crop performance, or of what the purpose of such a process might be. Part of the improvements that have taken place in modern times were within the reach of the Roman economy but there were limitations.³³⁹ The experimental approach that led to scientific developments never emerged.³⁴⁰ On the other hand grain is not the only product of these crops. Substantial amounts of straw are produced – which by weight is about two to three times the amount of wheat.³⁴¹ The improvements to yield would have reduced the amount of straw, which is needed. This was also a limiting factor.

6.4 Conclusions

As Trimalchio's freedman guest says: I know no geometry, fancy criticism or any such meaningless rubbish, but I do know my capital letters and I can work out percentages in weights and measures and currency.³⁴² There was a wide capability to undertake the ranges of work and specialities because of job specific literacy skills. The language of finance, the

³³⁵ Sallares (1991) 389.

³³⁶ Spurr (1986) 41.

³³⁷ Spurr (1986) 41-2.

³³⁸ Pliny the Elder, *Naturalis Historia* 18.55.

³³⁹ Improved husbandry through the withholding of nitrogen at early stages of growth to reduce stalk length and its introduction at later stages to promote growth of the grain, can offset the disadvantage of lodging to some extent and improve yield. Other changes were possible. Improved seed beds would have reduced losses and thereby increased yields. The Roman practice was to scatter seed rather than to drill into the ground – drilling ensures even spacing and maximises grain-to-soil contact. Conversation with P. Mulcahy.

³⁴⁰ But see Varro, *Res Rusticae* 1.18.7-8, discussed in Chapter 2, which describes a rich experimental approach.

³⁴¹ Gregg (1988) 70-5.

³⁴² Petronius, *Satyricon* 58.

concepts used and the understanding of the relationship between interest and time in the valuation of assets was sophisticated.

The deployment of knowledge and the sponsorship of projects such as the canal at Nicomedia and the construction of the ports at Portus were considerable enhancements to the infrastructure of the empire and greatly facilitated trade and promoted growth.³⁴³

The process of intellectual exploration and curiosity that emerged in Europe starting in the sixteenth century and the creation of new languages such as the mathematical language of calculus did not happen. The unintended consequences of that development, the creation of new material goods which generated potential for further economic growth, were not available. Substantial improvements in productivity in the area of agriculture, for example, were within reach but were not seen.

³⁴³ In this respect I disagree with Rostovtzeff's very negative view of Trajan's contribution to the economy by. Rostovtzeff (1957) 354-62.

CHAPTER 7: The money system, its weaknesses and the consequent limitations on growth

7.1 Introduction

The purpose of this chapter is to describe the monetary system that had developed by the end of the Roman Republic and to show how it changed through the first three centuries AD. Money had become essential to the functioning of significant parts of this economy and determined the nature of many of the economic relationships. When the monetary system collapsed those parts of the economy which depended on money were undermined.

7.2 Monetisation

The view that this economy was highly monetised and integrated has developed over time. Writing in 1970 Crawford took the view that coins were required at all social levels but that their use as a means of exchange was largely limited to the cities.³⁴⁴ Crawford in 1974 using an analysis of coin hoards demonstrated a very substantial growth in the money supply in the period 157 BC to 57 BC.³⁴⁵ Rathbone in his analysis of the Heroninos archive in 1991 found that money was essential to the functioning of large estates in third century Egypt and was used in the purchase and sale of goods and in the remuneration of workers.³⁴⁶ In 1992 Howgego argued that coin was used not only in the cities but also in rural areas and across all provinces regardless of the level of economic development. Coin was not the only means of exchange, however, and wheat played a role in the payment system being used in the settlement of some taxes and rents, for example.³⁴⁷ From an analysis of the proportions of coins minted under different Emperors and found in a selection of regions Hopkins concluded in 1980 that the economy was monetarily integrated with coins circulating freely through the empire.³⁴⁸ Kay in 2014 put this increase in the context of the increased supply of metal because of the proceeds of conquest and the acquisition of the Spanish mines. He also noted the increased use of the Roman denarius outside Italy but within the areas controlled by

³⁴⁴ Crawford (1970) 42-5.

³⁴⁵ Crawford (1974) 694-707. The work was subsequently criticised but these criticisms are dealt with by Lockyear (1999).

³⁴⁶ Rathbone (1991) 318-30.

³⁴⁷ Howgego (1992) 30. It is likely that the level of use of money as a means of exchange was lower outside the cities. My grandparents, who were farmers in Ireland, needed no money for milk, eggs, bacon and most of the fruit and vegetables they consumed. Nor did they need it to feed their horse. Their friends in the local town needed money for almost everything they consumed.

³⁴⁸ Hopkins (1980) 113.

Rome.³⁴⁹ Through this period the use of money had increased, the area in which coin was used became more extended and more of the transactions in the monetised areas were conducted with coin.³⁵⁰

Money is not essential for economic relationships. Near Eastern economies in the third millennium BC, for example, functioned without money and relied instead on centralised bureaucratic structures and the exchange of individual transaction receipts rather than the exchange of coin.³⁵¹ Neolithic communities relied on internal transfer of goods and services without any intermediation. The use of money is important in the development of economies because of the richness it brings to the structure of economic relationships. ‘The use of money makes it possible to obtain goods which are separated from those offered in exchange for them in space, in time, in respects to the persons involved, and, what is very important, in respect to the quantity on each side of the transaction. This results in a tremendous extension of the area of possible exchange relationships.’³⁵² Exchange relationships can take place between individuals who have nothing to offer each other than coin. The demands of individuals rather than just the state drive economic productivity. Saving is facilitated and individuals can produce goods and services now in order to consume later.

The economic relationship which had emerged through the development of monetisation in the Roman economy is illustrated by the diversity of the urban occupations described in Chapter 3. This level of private endeavour required a money system in order to function. The teacher of rhetoric acting independently of one master and taking on several pupils sought out people for whom he could work and did not require that they produced the goods that he needed. The tailor who rented his workshop at Rome collected coins from customers which he used to meet his obligations to his landlord, pay for his raw materials in the market and buy food. Long distance trade, described, in Chapter 4, was also dependent on the existence of a supply of money which could be accessed from sale proceeds, accumulated savings of the merchant and by borrowing from the accumulated wealth of others. All of these relationships depended on the existence of a sufficient supply of circulation money which was accepted in exchange for goods and services. Destruction of currency or the undermining of confidence in its acceptability worked to destroy those relationships. This level of

³⁴⁹ Kay (2014) 93-105.

³⁵⁰ It is possible that this increase in monetisation facilitated the process of urbanisation.

³⁵¹ See Nissen, Damerow and England (1993) for a detailed description of such a system.

³⁵² Weber (1978) 80.

economic activity was made possible by the growth in the money supply which had started in the middle of the second century BC.

Control of the currency, in particular control over the change of bullion into coin, is a source of economic power. In the eighteenth century the Nawab of Bengal and the East India Company concluded a treaty which illustrates of the importance of the power to mint coins. This treaty included among other provisions two currency concessions by the Nawab. The first concession was to place the rupees minted in Madras by the East India Company at a par with those minted by the Bengal government and the second was to allow the company to use the Bengal mint to change bullion into coinage. The Nawab ignored these concessions and doubled the discount placed on the Company's Madras minted coins in order to force the Company to import bullion rather than coin. By refusing to accept bullion into the Bengal mint he acquired a stranglehold over the Company's business in his territory. The attempts by the Company to borrow money were thwarted since the only source of loans in the territory were the Nawab and the Hindu bankers who controlled the mint.³⁵³ In essence the Nawab of Bengal was able to control trade within Bengal, and to exercise control over those who wanted to trade there, through his control of the mint because the coins produced by that mint were needed for trade. I will argue that the control exercised over the minting of the currency was critical to the stability of the monetary system of imperial Rome.

7.3 The dynamics of the metallic currency

Basically speaking, the currency consisted of the gold aureus, the silver denarius and the bronze as. The exchange rates between these coins was fixed by Augustus as one aureus to 25 denarii and one denarius to 16 asses.³⁵⁴ The currency was designed so that the gold coins would be used for higher value transaction – such as the purchase of capital goods and more valuable consumer durables.³⁵⁵ These high value coins tended to travel very fast.³⁵⁶ Gold coins were, at least until the late third century, used routinely in transactions.³⁵⁷ The silver coins were used for routine purchases of consumer goods and for the daily wages of

³⁵³ Keay (1993) 232-3.

³⁵⁴ The relationships are broadly similar to those that applied in pre-decimal English coinage when the factors were 20 and 12. The coins in both systems were designed for broadly similar purposes.

³⁵⁵ The relationship between the prices of one gram of gold to one gram of silver was about 15:1 until the middle of the 19th century. This is also the relative abundance of the two metals. The design element is to fix the weight of the aureus at twice the weight of the denarius.

³⁵⁶ Crawford (1970) 43.

³⁵⁷ Communication from M. Crawford indicates that hoards of gold before the late third century regularly show normal degrees of wear on older pieces.

labourers.³⁵⁸ The copper coins are used for trivial purchases and for fractional prices. Through our period, the first to the third centuries AD, the silver coins went through a process of debasement.³⁵⁹ This was not the case for the gold coins, although the weight of these coins was decreased.

Some uniformity was introduced into the currency as the areas under Roman control expanded and in time the mint was centralised.³⁶⁰ Nevertheless separate coinages did exist and Egypt, for example, continued to have its own coinage system.³⁶¹ From its annexation in 30 B.C. the Egyptian economy remained as a separate currency area. The unit used in Egypt was the Alexandrian tetradrachm, which was broadly equivalent to the denarius although the denarius was forbidden to be used as a unit of currency in the province. For most of our period the lower value coins used in the eastern half of the Empire were produced there.³⁶² Some Greek cities produced their own silver coins and most produced their own bronze coins.³⁶³

Although Augustus had set exchange relationships between the different denominations these were not always observed, at least as far as the bronze coinage is concerned. An inscription records an attempt by bankers in the time of Hadrian to convert coinages at rates different to the official rate and traders, retailers and fishmongers petitioned against this and tried to get the state to force the use of the official rate.³⁶⁴ It is unlikely that this was an isolated case. The exchange rate of denarius to as varied from 16 to 24, with the number increasing over time.³⁶⁵

Lo Cascio argued that an official policy existed to maintain the relationship between the gold, silver and copper denominations through manipulation of the metal content and face value.³⁶⁶

³⁵⁸ A.H.M. Jones (1953) sees these coins as being used for most transactions.

³⁵⁹ The debasement of the denarius and the Alexandrian tetradrachm follow similar patterns. Rathbone (1996a) 327.

³⁶⁰ Howgego (1992) 5 describes the withdrawal of most non Roman silver and gold coinages noting that the main gold coinages of Macedon, Carthage, Ptolemaic Egypt and Gaul were no longer in circulation by 46 BC. In the later part of the first century AD Vespasian closed the local mints and centralised the imperial and senatorial mints at Rome into one which produced the coins used in the West and to an extent those needed in the East. The centralised production of coins continued until AD 193 when minting resumed at Antioch. Mattingley and Sydenham (1992) 7. Almost all of the gold coinage in the Empire, the dominant silver coinage and the base metal coinage for the West were decentralised by AD 250. Howgego (1994) 6.

³⁶¹ Rathbone (1996b) 317.

³⁶² Howgego (1994) 6.

³⁶³ Through the first century AD the denarius became the only or the dominant silver coinage, replacing local coinage, in most areas of the empire. East of Greece and Cyrenaica a few local coinages survived. Burnett, Amandry and Ripollès (1992) 1 6. Mattingley and Sydenham (1992) 12 reference the circulation in the East of local silver coins in competition with the denarius and note the local production of copper coins.

³⁶⁴ Crawford (1970) 42.

³⁶⁵ Crawford (1970) 43.

³⁶⁶ Lo Cascio (1981). Wassink (1981) 476 follows this view.

Rathbone argued against the view that there is management to keep the relationship right and the denarius relative to the aureus had a face value which was more than its metal value.³⁶⁷

The relationship which was established between the aureus and the denarius may have been designed to prevent arbitrage. The metal relationship between the coins was 12.5:1. This relationship seems to have held at least until the time of Caracalla (AD 194-217), despite the changes in the silver content of the denarius, but had been abandoned by the time of Phillip the Arab (AD 244-249).³⁶⁸ This fixed relationship, particularly between the silver and gold coins, is important for commerce. If the fixed relationship is widely accepted and is trusted to apply in the future then individuals will freely exchange small numbers of coins of high value for large numbers of coins of low value and conversely. Exchange between individuals who produce goods and services which have different values is facilitated and price negotiation is limited to the price of these goods and services and not to the relative values of the coins used in the transaction. The present and future relative value is taken as a given.

The conventional view among economists is that bimetallic systems, let alone tri-metallic systems, are unstable.³⁶⁹ The fundamental difficulty of the bimetallic system is that an instability arises from the fact that the coinage is based on fixed exchange rates while the relative market value of the two metals are variable. For example, an increase in the supply of gold because of the capture of substantial gold treasures will drive the bullion price of gold down relative to silver. The heavy production of silver coins for whatever reason will reduce the supply of silver on the market and increase the silver price relative to that of gold. The existence of fixed exchange rates for the bimetallic coins in the context of variable exchange rates for the metals in theory should encourage constant flows of the coins from where they are undervalued to where they are more valued. And since the market value of the coins will seldom be precisely the same as the fixed exchange rate the question is not whether there is an arbitrage opportunity but rather in which direction it lies.

Scenario A and Scenario B are hypothetical scenarios which might plausibly arise. In each of these scenarios there is a possible strategy that could be applied and which generates substantial profits.³⁷⁰ Any number of different scenarios are possible and for each of those

³⁶⁷ Rathbone (1996b) 319.

³⁶⁸ Crawford (1975) 566, 569.

³⁶⁹ By bimetallic I mean systems in which the state issues coins in two metals and sets the rate at which these coins are to be exchanged with each other. Friedman (1990) discusses the arguments. Friedman (1990) 102 gives the requirements for such a system to work as the widespread use of deposits and reduced use of coin.

³⁷⁰ I use the Augustan minting standards. The aureus was produced at the rate of 40 coins per pound of gold. The denarius was produced at the rate of 84 per pound of silver.

scenarios a profit opportunity is available. What varies is whether the strategy to be applied involves the destruction of denarii or aurei and the extent of the profit available. But in any case the wider the gap between the relative market prices of gold and silver and the relative values placed on gold and silver coins in the official exchange rate then the greater the profit opportunity. This is an inherent weakness in a metallic system of money.³⁷¹

³⁷¹ For example when Domitian increased the purity of the denarius to 98% then the new coins disappeared quickly and the 90% coins continued to be used. It is easy for the moneychangers to detect such a difference and exploit it. Discussion with K. Lockyear. Trimalchio expresses his admiration for the man who could tell good from bad coins. “What profession do you consider the most difficult after that of letter? I myself imagine it is ... that of the doctor or of the bank teller ... the doctor because he spots what’s beneath the skin ... and the teller because he spots the copper beneath the silver. Petronius *Satyricon* 56. Translated by Walsh.

Table 7.1 Illustration of arbitrage opportunities in a metallic currency

What one gramme of gold trades for on the market	Situation	Possible strategy	Consequence	Consequence
Scenario A 15 grammes silver	Relative shortage of gold	Exchange 375 denarii for 15 aurei Melt 15 aurei and exchange for 473 denarii worth of silver Profit margin 28%	Gold tends to disappear from coinage or from circulation	Reluctance to exchange aurei for denarii
Scenario B 10 grammes of silver	Relative shortage of silver	Exchange 10 aurei for 250 denarii Melt 250 denarii and exchange for 12 aurei worth of gold Profit margin 17%	Silver tends to disappear from coinage or from circulation	Reluctance to exchange denarii for aurei

These arbitrage opportunities involve destruction of currency and put pressure on the official exchange rate. The exploitation of this very profitable arbitrage opportunity, therefore, serves to reduce the amounts of money available for transactions. A relative shortage of gold in the market would make transactions in capital goods more difficult and a relative shortage of silver in the market would makes transaction in consumer goods more difficult. Exploitation of the arbitrage opportunity also adds uncertainty to the rate at which an aureus will be exchanged for denarii. Both of these consequences, reduction in money available and uncertainty as to the rate at which coins are exchanged, are harmful to commerce. Both of the strategies outlined above produce high profit margins. The proceeds of these strategies are quantities of precious metals. Where these quantities of precious metals can be brought to a

mint and converted to coin which can be used in commercial transactions then the pressure on the currency becomes acute. Even in the absence of a mint which will reconvert the bullion extracted back into coin monetisation remains possible since the metals are valuable, can be used in manufacture and the product sold.³⁷²

Counteracting this pressure on the coinage was the need that individuals had for coins to acquire the necessities of life, for tax and for the conduct of business. Melting down of coins and the exchange of metals will only occur when those coins are excess to the individual's need for coins. Where money is limited relative to the demand for it the pressure is reduced, although individuals with excess coins will continue to have the incentive to destroy money. In times of increasing trade, as for example in the first and second centuries AD, there is increasing demand for coins and in times of reducing trade, as for example in the third century, there is reduced demand. Equally, there are costs involved in applying the strategies described in Table 7.2 and where the difference between the rates at which metals exchange in the coinage and in the market place are small then costs may outweigh the gains from currency destruction.

The bimetallic problem was first solved in 1816 when the UK enacted legislation which for the first time established a stable coinage which used two metals. The measures contained within that legislation were necessitated by the fact that gold and silver coins were leaving the country and commercial transactions were becoming difficult. The pressure was particularly acute because of the easy access between England and Holland which operated two separate currencies. Indeed to overcome the difficulties caused by the lack of coin some merchants were manufacturing their own tokens to facilitate trade. The first measure was to limit the extent to which the silver coinage was legal tender. The limit was set at £2 or 40 shillings. The second rule was that while a troy pound of silver could be coined into 66 shillings it would be sold by the state for 62 shillings. The effect of these two rules was to generate a stable currency free from the arbitrage difficulties mentioned above and was 'one of the chief causes of England's long continued dominance in the field of international finance'.³⁷³ In essence what had been discovered was a mechanism to make one of the coins subsidiary to the other which then became the standard.

³⁷² A further arbitrage risk exists since coins wear down with use and the metallic content of old coins will be less than that of new coins issued to the same standards. Butcher and Ponting (2014) 37.

³⁷³ Scott (1910) 315-8.

The first rule removed the risk that coins could be acquired cheaply when the price of silver fell and then dumped on merchants. It also tended to limit hoarding. An essential development in the 1816 legislation was to find the legal tender limit that worked.³⁷⁴ The first rule, therefore deals with, amongst others, the problem of what happens when the price of silver drops.

The second rule deals with the problem of what happens if the price of silver rises. Melting down silver shillings made no economic sense because while 66 shillings will produce a troy pound of silver, the state will give you 1.06 troy pounds of silver for 66 shillings. The private arbitrage opportunity has been shut off.

It was these two rules, together with rule introduced in 1774 that made gold coins legal tender in unlimited amounts, provided the gold coins were of full weight that made for a stable currency and paved the way for the creation of fiat money. When in 1920 the silver content of the shilling was reduced by 50% the debasement had no inflationary effect and did not disturb confidence in the coinage. The shilling had by then become a universally desirable commodity, a necessary condition for coinage, and indeed its face value had become twice its intrinsic metal value.³⁷⁵

Metallic money systems with fixed exchange rates between the coins are unstable. There were constant incentives for the private sector to destroy currency and the relationship between the different coins within the system was unstable. The simple ideas necessary to achieve currency stability and to start the process to the creation of a fiat currency only emerged by the early nineteenth century in England and currencies tended to remain valued largely, but not entirely, for metallic content. Nevertheless, for long periods the Roman imperial coinage was a stable currency, did not collapse and indeed continued to be used despite being debased. Control of the currency by restricting minting rights within the Empire together with the need for coin to pay taxes, as well as the commercial need for coin, which provided demand for the currency all acted against the pressures towards destruction of the currency. In any case, restriction on the use of other coinages and the applicability of those controls over a large geographic area limited the extent to which the inherent weakness of a

³⁷⁴ The earlier legislation of 1774 had set the limit at £40 which was too high and proved ineffective.

³⁷⁵ Silver content 5.7 grams. Price of silver 0 .65 USD per troy ounce. 1 USD was 0.255 £ then. 1 troy ounce is 31.1 grams.

metallic currency manifested itself. Weakening of those controls was destructive to the currency and to trade.

7.4 Relationship between the silver content of the denarius and inflation.

This subsection considers two particular features of the metallic currency which affect price. First there is a relationship between the price of silver and the silver content of the currency. Secondly there is a relationship between the price of silver and the price of other commodities. These factors act as restraints on price inflation.

Let us consider the relationship between the denarius and the silver price. The Augustan denarius contains 3.8 grams of silver.³⁷⁶ The silver price will not rise much above 26 denarii for 100 grams of silver because melting denarii down then becomes a cheaper way to acquire silver. If the price of silver is 30 denarii, say, for 100 grams of silver you are better off melting the 30 denarii down because you get 114 grams of silver. There is some cost and inconvenience to the process of melting down coin but with any significant divergence from the price of 26 denarii there is a profitable business opportunity for those individuals with the necessary technology. Making the destruction of the coinage illegal is an important but not insuperable limiting factor.

Similarly, if the price of silver falls to say 20 denarii for 100 grams of silver then there may be reluctance for those who have 100 grams of silver to sell for 20 denarii and acquire coins with 76 grams of silver. The underlying economic demand for coin, however, acts as a countervailing pressure. Coin was essential to the economic relationships that had developed and there was a need for coin to buy basic goods and services. The price of silver could therefore fall below 26 denarii for 100 grams simply because people needed denarii to buy wheat. However if the supply of coin exceeds the demand for coin then this pressure is weakened. Equally, if there is a mint which will accept bullion and give coin in exchange then it might be profitable to acquire 100 grams of silver for 20 denarii and have this coined into 26 denarii, less the costs of minting. This last strategy was in any case available to those who controlled the mint.

There is therefore pressure on the price of silver to equate to the silver content of the coin. The tendency for the price of silver to reflect the silver content of the coinage acts as a restraint on inflation. Provided that the relative supply and demand for silver and wheat, for

³⁷⁶ The minting target was 84 coins per pound of silver.

example, do not change then the rate at which they can be exchanged for each other will stay constant. If the silver content of the coin stays the same then the price of wheat will remain constant. In a later section of this chapter I will show that while the silver content of the coinage was very broadly stable, which is to say over the first two centuries AD, there was limited price inflation.³⁷⁷

7.5 Money supply-features of this money system

In modern economies money is created out of thin air by central bank through asset purchase. Money is further created in the private banking system through the multiplier effect.³⁷⁸ The limiting constraints, on money creation in the private banking system are the reserve ratio and the knowledge on the part of the authorities that excessive money creation destroys the system through hyperinflation. A significant difference between the Roman economy and a modern economy is that in a modern economy money is created at will but in the Roman economy state ownership of precious metal is a prerequisite.

7.5.1 State money creation

Thus there are two ways for the state to create money in this system, both of which operate through the mint:

First Model:

State acquires reserves of monetisable assets through conquest, expropriation or some such. We do not count here its own coins which it receives back through taxation (and some appropriation) since this is not new money. Monetisable assets include bullion, reserves in mines, assets in temples, overseas assets appropriated and overseas currency not accepted in the economy.

State converts some of its reserves into coin (M)

State issues money M.

Second Model:

State destroys money collected through taxation or otherwise. (Destroy N)

³⁷⁷ Butcher and Ponting (2014) 40 describe an alternative view which is that the state's acceptance of coin as payment for taxes maintained confidence. I do not believe that this confidence would have been sufficient to dampen price pressures.

³⁷⁸ See Whelan (2012) for a description of modern money creation.

State increases reserves by the amount of valuable metal extracted.

State creates new money with a lower metal content (Create M)

State issues money M

In the first model new money M is created. In the second model new money of M-N is created. Both models of state money creation have their limitations. The first model requires that there be resources available, output from mines or treasures taken from conquered territories, for example, sufficient to create new money. Under the Principate, when the rate at which new territories was being acquired reduced, the ability to increase reserves was more restricted but nevertheless there were opportunities. When Tiberius brought the gold and the silver mines under Imperial control he increased the state's capacity to create money.³⁷⁹ The state's capacity to create currency was increased by treasures acquired from the sack of Jerusalem in AD 70, the conquest of Dacia in AD 106 and the capture of the royal treasures of the Parthians in AD 190. The abandonment of the Dacian mines in AD 167, however, reduced the capacity to create money.

Silver production may also have reduced in the period. De Callatây, drawing on work produced by Hong and others, summarised the development of aggregate silver production in the northern hemisphere. Lead is a significant by-product of the process of extracting silver from ore and can be 300 times the amount of silver produced. Part of that lead escapes into the atmosphere and part of that is then deposited on the Greenland ice sheet. Analysis of the lead content of samples extracted from the ice sheet can be used to generate a history of lead – and therefore silver – production. This analysis indicates a continuing increase in production until about the end of the first century AD. Production of silver then declined and it was not until the eighteenth century that world production was at the level seen at the height of the Roman Empire.³⁸⁰ There is evidence, at least for the Iberian mines which were the principal source of precious metals mined in the period of the Principate that large scale mining may have peaked in the first and second centuries AD.³⁸¹ These both suggest that the flow of metal from mining into the reserves of the state may have declined during the second

³⁷⁹ Confiscations under Tiberius, while they served to increase the state's reserves, reduced the money supply.

³⁸⁰ de Callatây (2005) and Hong, Candelone, Patterson and Boutron (1994). Note the relevant data is presented in graphical form in these papers and the dates for silver production decrease and increase used here have been read off the graphs and have a degree of approximation.

³⁸¹ Edmondson (1989). Harl (1996) 81. Jones (1980) indicates a very sharp decline in activity in the silver mines at Rio Tinto at the end of the second century AD and attributes this to the loss of Roman control over Southern Spain in the 170's AD.

century and that creation of money by the first model may have subsequently become more difficult.

Both models of money creation have their limitation and the second model of money creation described above has the limitation that it can only be repeated for so long, because in time only limited amounts of precious metal will be contained in the coins being collected as tax payments and there is a limited extent to which the coinage can be further debased.³⁸² The following graph presents an index of the rate at which the coinage was being debased. The index has been calculated by dividing the numerical value of the silver content of the denarius in AD 60 by the numerical value of the silver content of the denarius at each subsequent point in time. So, for example, the silver content of the denarius in AD 64 was 2.97 grams and was 1.46 grams in AD 241. The index for the time of Augustus is then 1.0 and for AD 241 is 2.03. See Appendix 4 for discussion of the values used in constructing this graph. Although there had been changes in the first and second centuries in the silver content of the denarius, these were minor. Changes in the third century were of a different order of magnitude. In about AD 250 you needed to collect over seven newly minted denarii to have as much silver as was contained in one denarius minted under Augustus. In about AD 270 you needed to collect seventy five newly minted denarii to have as much silver content as was contained in one denarius issued under Augustus. This debasement had been attributed to various factors. Walker in 1978 attributed the debasement that had occurred in the period AD 64 to AD 253 to periodically occurring high levels of state expenditure generally related to warfare, imperial anniversaries or 'extravagances' and to avoid the need to increase taxation.³⁸³ Wassink in 1990 attributed the debasement of AD 238 to the fact that there was not enough money to pay the army.³⁸⁴ In 1992 Howgego attributed the debasement to the shortage of gold and silver.³⁸⁵ Rathbone in 1996 took the view that debasement was a government reaction to a rise in prices.³⁸⁶

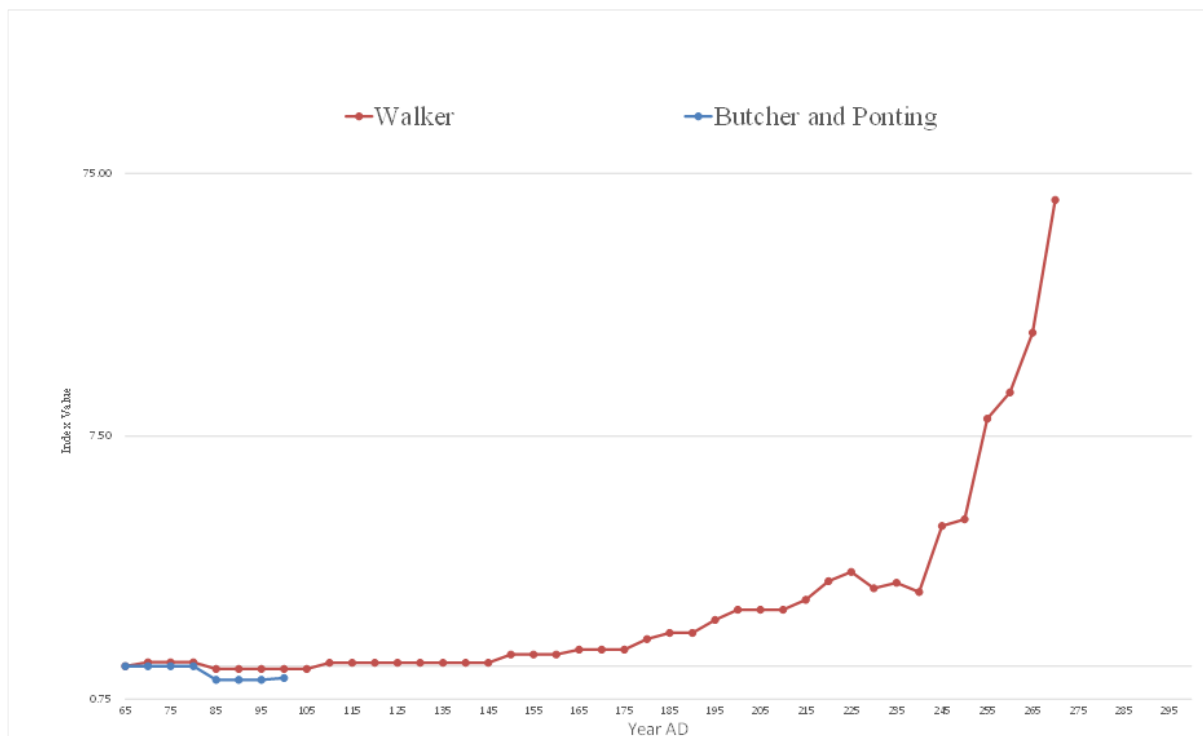
³⁸² For example if new coins are debased by reducing silver content by 10% each year and if there is a very high propensity for the coins returned to the state through the payment of taxes to be those of poorer quality then after twenty years at this level of debasement the silver content of the coins being minted is $(0.9)^{20}$ which is 20%. In practice the silver content will be greater than this because at least some taxpayers will return better quality coins either because of limited resources from which to select or because of an inability to discern the relative qualities of the coins. Actions by the state to select the quality of the coin which is suitable for the payment of taxes in order to protect the flow of precious metal to the mint would undermine the confidence in the coins being issued.

³⁸³ Walker (1978) 109, 138.

³⁸⁴ Wassink (1991) 482.

³⁸⁵ Howgego (1992) 8.

³⁸⁶ Rathbone (1996a) 334.

Figure 7.1 Debasement of the coinage

I place this debasement in the context of the state's varying need over time to create money and its varying ability to do so. Money creation by the second method become increasingly difficult from the middle of the second century AD because of the limited silver content in the coins. The relationship between this index and inflation is discussed later.

The last stage in the creation of money is its issuance by the state. In modern economies money is issued by central banks when they purchase assets either by issuing currency or increasing the reserves held to the credit of private banks at the central bank. In the economy of imperial Rome the state issued money through the state purchase of goods and services. Money was also issued through the payment of tribute to external parties and other non-purchase transactions, including the expropriation of state coin by private individuals.³⁸⁷ Taxation used to purchase goods and services is recycled money, not money created. Although there has been some argument that the authorities were conscious of the need to manage the currency for general economic reasons, I take the view that issuance of money was expenditure related and that new money created was the excess of total expenditure over

³⁸⁷ The coining, for private individuals, of precious metals is another way to issue money. This does not appear to have happened at Rome before the fifth century AD. Howgego (1990) 19.

money collected as taxation and recycled out as state expenditure.³⁸⁸ I take the basic model of the final stage of money creation in the economy of imperial Rome to be:

$$\text{New Money created} = \text{Expenditure} - \text{Taxation} \quad \text{Formula 7.1}$$

and since

$$\text{Fiscal Deficit} = \text{Expenditure} - \text{Taxation} \quad \text{Formula 7.2}$$

it follows that

$$\text{New Money created} = \text{Fiscal Deficit} \quad \text{Formula 7.3}$$

In the imperial Roman system there is a relationship, therefore, between the amount of money being created and the fiscal deficit.³⁸⁹ Capital projects such as the draining of the Fucine lake in AD 41 or the heavy expenditure of Nero, the extensive issuing of coins by Septimius Severus at the end of the second century and the regular expenditure on the army, to the extent that they exceeded taxation, were a constant source of money creation. The important common feature however is that new money issued in this structure is equal to the fiscal deficit. The money creation process relied essentially on the state being able to fund itself from sources other than taxation and from the state spending rather than hoarding its reserves. The reminting of coins, for example, the taking into the mint of coins issued under a previous emperor and the reissuance of their metal content as coins struck with different image does not amount to money creation except to the extent that the coins were debased.

Hopkins, writing in 1980, provided a quantification of the level of money creation. He categorised coin finds for the period AD 40 to AD 260 by broad area of find and then generated for each area an index which enabled the relative volumes of coins minted at each date to be compared.³⁹⁰ The indices for each of the territories have the same general shape which Hopkins took to indicate a high level of monetary integration and I agree with this

³⁸⁸ See Lo Cascio (1981) who argues that the authorities noticed the lack of currency and acted. Howgego (1990) takes the view that coins were produced for reasons other than expenditure. Crawford (1970) 46 was of the contrary view that coins were only issued to make payments.

³⁸⁹ Note that this is gross money created. The net increase in money supply will typically be lower because of factors to be discussed shortly.

³⁹⁰ Hopkins (1980) 112-116. The index was calculated as follows. Coins could be identified to the reign in which they were minted by examination of the image of the emperor on the coin. He then divided the numbers of coins associated with each emperor by the length of his reign to derive the numbers of coins issued in each year of his reign. This assumes uniform issuance which is a satisfactory assumption. He then took the average for the period AD 96 to AD 180 as the base for his index. The indices have some limitations. They include no quantification of money destruction whether simple loss of coin or melting down. They do not allow for the effect of reminting of coins. These limitations do not undermine the general conclusions drawn above. See Howgego (1992) 2-4 for some discussion on the difficulties of quantification. Lockyear (1999) is also relevant.

conclusion. These indices show uniform and very dramatic increases in the level of minting of new coins in about AD 180 to AD 200. Hopkins suggested that this increase in minting could be associated with an increase in the money supply and a consequent change in the level of prices and this aspect is dealt with later in this chapter. I believe that a more fundamental point is that the dramatic increase in coins minted at the end of the second century suggests a dramatic increase in the level of state expenditure. The next chapter discusses the cause of that increase in expenditure. The level of debasement seen from the end of the second century AD and shown in Figure 7.1 above implies that the state had to rely on the second method for creating money rather than solely on the first method described earlier. In other words, the level of reserves of the state may no longer have been sufficient to fund the state deficit. I believe that the principal cause of this dislocation was the extent of the increase in state expenditure which occurred at the end of the second century. The indices also show, as Hopkins notes, a significant increase in minting in the reigns of Vespasian and Titus, AD 69 to AD 81, although considerably less dramatic than the increase shown for the end of the second century AD.

The pressures that tended towards money destruction have already been noted. These pressures were further increased by the issue of debased coinages, which tends to cause the earlier issuances with higher metal content to be either hoarded or destroyed, but in any case to be removed from circulation. Money was also being destroyed because coins were lost, but this was most likely minor compared with the other influences.³⁹¹ The loss of money through trade may have been a significant factor.³⁹² The money created by the excess of expenditure over taxation and evident in the levels of new coins minted is offset by these factors. Indeed since there is a continuing reduction in money supply simply because of the loss of the coin supply money supply only stays constant where new money created exceeds that loss.³⁹³ Any

³⁹¹ See Lockyear (1999) 240 for decay rate.

³⁹² Turner (1989) 6-12, 16, 28 analyses coin hoards from India and sees these as indicating a significant increase in trade under Augustus. The imports from India were gemstones and other expensive items. The West had little more than gold and silver to offer in exchange. The exported coins were treated as bullion rather than used as coin by the Indians. Most coins from the Julio-Claudian period found there are silver and the finds with later dates are exclusively gold. See also Rathbone (2000) for discussion of a trading venture with India. Wassink (1991) mentions disappearance of silver across the borders 'mainly to free Germania where silver was the commodity in which wealth was secured and hoarded.' Howgego (1992) 5-6 discusses loss of coin through trade. A similar situation faced England in the seventeenth century when concerns were raised about the export by the East India Company of precious metals to India in exchange for luxury consumer items, including manufactured goods. A director of the company countered that so long as the values of the goods imported into England exceeded the value of the goods exported the trade was beneficial to the nation. See Keay (1993) 119.

³⁹³ There is a difference with modern economies where most of the currency in circulation is paper based. Loss of money can be offset by printing at limited cost and no inflation risk. The loss is to the individual who holds the paper and generates a gain for the state.

coins lost from circulation within the empire because of trading beyond its borders represent a reduction in money supply. Debasement of the coinage, particularly where this is evident to the user, will tend to lead to withdrawal of better quality coins from circulation. The increase in minting noted by Hopkins and described above was at least partly associated with a debasement of the coinage and will not have fully reflected itself in an increase in money supply. Money supply was under continuing downward pressure and was only sustained if state expenditure exceeded taxation, which in turn required that the state was able to create money to fund the difference. Offsetting this pressure is the additional money created through private banking which I now describe.

7.5.2 Private money creation

Harris, writing in 2006, argued that the money supply did not consist just of coins. The principal arguments for that view were that lending of money was widespread, that large transactions could not conveniently be settled in coin and were not settled in bullion, that deposits with bankers were transferable and that overseas trade transactions were not financed by coin since there is limited evidence of coin in shipwreck. He also noted that payments could be effected between two parties at different locations through agents with whom they held balances.³⁹⁴

Rathbone and Temin (2008) described the banking operations of the imperial Roman economy. These operations are described in terms of the activities of individuals which because of similarities to the activities of modern banks allows us to call those individuals bankers. They accepted money as deposit and generally paid interest on those deposits and they made payments on behalf of depositors. They advanced loans out of the funds that they had collected as deposits. They carried out other functions such as guaranteeing payments, financing purchases at auction and changing money. Banks were required to keep records of money taken in and paid out and to make relevant entries in their books available in legal cases. Banks could be formed as partnerships in the form of a *societas*, as described earlier for trading partnerships, or be established as an agency of an individual with the agent, slave or freedman of the individual, acting under defined authority particular to that agency. The banking operations took place through meetings between banker and customer generally in the public areas where most business was carried out rather than in a building specifically used by the individual banker. Banking operations were widespread and many towns had

³⁹⁴ Harris (2006) 1-24.

more than one banker. Some at least of these banks had more than one branch or at least had business relationships with banks at other locations.³⁹⁵

This banking structure, which operated through business meetings in public places and which was embedded in legal structures, was sufficient for the creation of money beyond that in circulation as coin. If a banker with deposits of HS1,000 lends these deposits to an individual who uses that money to buy goods from a third party who then in turn deposits the funds with the original banker then that banker holds deposits of HS 2,000 of which HS 1,000 are in coin and HS 1,000 are in loan assets. The amount of money in the system has increased and HS 1,000 has been created.³⁹⁶ As noted above, most towns had bankers and safe keeping of funds and interest earned on deposit were incentives to place money with such bankers although there were alternative ways to achieve these ends including loans directly contracted between the individual with surplus funds and the borrower. The settlement of payments between individuals who each had dealings with the same banker and therefore the use of deposit as money is possible under this system.

This business model, which requires the handing over of funds in the expectation that these funds will be returned when due requires belief on the part of the depositor that the banker will do just that and depends on the ability of the banker to provide funds when needed. The system contains the same solvency and liquidity risks as modern systems. Poor lending by the banker can lead to loss on loans and that or holding insufficient coin in reserve can lead to inability to meet calls for return of funds deposited.

The absence of coins or bullion in shipwreck suggests that trading operations may have been funded through the banking system and in any case funding through a network of bankers is more secure than sending coin on voyage subject to the risk of shipwreck or piracy.³⁹⁷ Indeed at its simplest where ships voyage between location A to location B carrying one set of goods from A to B and another set of goods from B to A then at both locations there are buyers and sellers who each have need for funds and expect to receive funds. The difficulty is that those who need funds at location A, to buy goods there, expect to receive funds at location B, when they sell goods there and those who need funds at location B expect to receive funds at

³⁹⁵ Ratbone and Temin (2008) 391-407. More informal sources of credit were also available. Holleran (2012) 52-3 gives an example of the sale of goods on credit.

³⁹⁶ See Kay (2014) 110-3 for a more detailed example. Kay (2014) 311-322 assumes that for 100 BC the level of deposits in Italy was 12% of the value of the denarii circulating there and that after the credit crisis of 80 BC it fell to 5%.

³⁹⁷ Communication from M. Crawford indicates that between AD 250 and AD 450 there are industrial quantities of coins in shipwrecks. This suggests a possible breakdown in the bank transfer mechanism from about AD 250.

location A. In theory two bankers, one at location A and the other at location B, could transact substantial amounts of business with limited transfer of coin and with their businesses supported by their system of keeping records and the belief that their obligations would be met exactly as due. As an example, trader A at Rome borrows HS 10,000 from a Roman banker and spends that HS 10,000 buying goods at Rome for sale in Alexandria. He sells his goods at Alexandria for HS 15,000 and deposits that money with a banker there. Trader B at Alexandria borrows HS 10,000 from that banker to buy goods which he sells for HS 15,000 in Rome and deposits that amount with the Roman banker. Both traders then have loans of HS 10,000 in their home territory and deposits of HS 15,000 overseas. In theory correspondence between the two bankers, or other documentary exchange, then allows trader A to access 15,000 in Rome and trader B to access 15,000 in Alexandria. Both can repay their loans and access their profits without having to transfer coin. Money is created through these transactions to the extent that they encourage the depositing and lending. This system depends for its continuing operation on the enforceability in both locations of contracts entered into in the other location. Such banking transfer mechanisms are not needed where a trader travels with his cargo from Rome to Alexandria, say, sells that cargo there and uses the sale proceeds to buy new goods to carry back to Rome for sale. Nevertheless, the banking structures that existed could be used to support very substantial amounts of trade without the need for corresponding amounts of coin and need not have been unduly restricted by changes in the money created by the state.

7.6 Inflation – the facts and the contributory factors

This section summarises some of the theories for the inflation that occurred during the Principate. Data on price changes is used to test these theories and also to highlight the many contributory factors to price change. The periods of price stability, of inflation and of hyperinflation are identified.

Both Suetonius and Dio were aware of the one of the causes of inflation – the effect of an increase in the money supply. When Augustus brought the abundant treasure of Egypt into Rome the supply of money increased to such an extent that the price of goods and land rose. The change in the money supply also had an effect on interest rates which fell from 12% to

4%.³⁹⁸ In essence what is happening here is that the increased supply of precious metals such as silver drives the price of silver down relative to land.³⁹⁹

Lendon argued in 1990 against the prevalent modern theories that inflation was caused by debasement or the increase in the money supply by attributing it to a loss of confidence in the currency. Instead he attributed inflation to psychological causes: the value of the coin was linked to the emperor whose face was on it and the death of that emperor could lead to the coins being less valued with a consequent increase in prices.⁴⁰⁰ Wassink in 1990 linked the high inflation seen in the third century to a supposed increase in money supply and the substandard nature of the coins issued rather than to a reduction in the supply of goods, which is another possible explanation. Rathbone introduced other causative factors arguing, for example, that the Antonine plague reduced the supply of labour and drove up its cost and that an increase in money supply need not lead to price increases since it might go along with an increase in monetisation. Moreover public perception might be a causative factor since fear of devaluation in the currency could lead to price increases.⁴⁰¹

There is a need to approach the question of inflation in this economy with an awareness of the many factors which can bear on an economy and which will exert different pressures at different times. While, for instance, a purely monetarist way of thinking about an economy may have been appropriate in the 1980s this is not always so. By way of example, the modern US economy is not run on the assumption of a clear connection between money supply and inflation precisely because of the need to allow for other factors. These other factors include changes in government spending, tax changes, shifts in confidence, changes in lending practices, the occurrence of natural disasters and slowdown in agricultural production. There is also uncertainty about the time lag between any change in money supply and its effect. There is, therefore, no immediate clear connection between changes in money supply and the economic response.⁴⁰² Equally, theories which fit modern money systems might not apply fully to the money system that prevailed in the imperial Roman economy. For example, destruction of money in circulation to extract precious metal content made economic sense in some circumstances in a way in which destruction of money in circulation in modern systems

³⁹⁸ Suetonius, *Augustus* 41. Dio 51.21.

³⁹⁹ These changes also imply that the treasure was either coin or converted into coin. Treasure which was not coined would not have had these effects.

⁴⁰⁰ Lendon (1990) 113-7.

⁴⁰¹ Rathbone (1996a).

⁴⁰² http://www.federalreserve.gov/pf/pdf/pf_complete.pdf 19-20. Retrieved 30 October 2015.

does not. Increases in money supply might be dampened by currency destruction rather than leading to price inflation or changes in velocity of circulation.⁴⁰³ Equally the silver content of the coinage might serve to dampen inflation, as described in Section 7.4, producing an effect not available in modern currencies.⁴⁰⁴ Additionally, an increase in money supply may be linked to an increase in monetisation. There are, therefore, limitations to the applicability in the monetary system of imperial Rome of the classic Fisher equation of $MV = PT$.⁴⁰⁵ the relationships implied by this formula may not hold when there is an increase in the supply of money. Reductions in the supply of money is, however, still likely to put pressure on transactions.

If the metallic theory is correct then we will expect to see a high correlation between changes in the prices of commodities and changes in the metallic content of the coinage. In what follows I test the theory by mapping the silver content of the denarius against the prices of some commodities and also against the average earnings of military personnel. Each graph shows an index of the silver content of the denarius over the period 60 AD through 270 AD displayed in Figure 7.1 After 270 AD the silver content of the denarius is so low that it hardly contributed to the desirability of the coinage. The graph shows the significant and continuing debasement that occurred over the period. I use these graphs to test the relationship between price changes and the debasement of the coinage and also to identify the periods of currency deterioration.⁴⁰⁶ I test this theory with data for wheat, wine and donkey prices drawn from Roman Egypt and with military pay data.

Figure 7.2 shows the progress of the prices at which the state bought wheat in Roman Egypt. The wheat price index values are circled. A wheat index value of 1.00 indicates that the price at which the state bought wheat is the same as it was at the beginning of the period. A wheat price index value of 3.00, for example, indicates that the price at which the state was acquiring wheat had trebled since the beginning of the period. Beyond minor variations and despite significant debasement in the coinage from the end of the first century AD the state

⁴⁰³ Crawford (1975) 591, for example, notes that currency was destroyed to make jewellery.

⁴⁰⁴ A further difference between the modern money system and that of imperial Rome is that in the modern system money is fungible. One dollar bill has the same value as another, regardless of when they were printed. Two denarii coins, with different silver content, are indistinguishable from each other.

⁴⁰⁵ Lockyear (1990) 242 notes his rejection of the classic relationship expressed in the Fisher equation.

⁴⁰⁶ I have calculated the index as follows: The figures for wheat are taken from Rathbone and van Reden (2015) and the figures for wine and donkeys are taken from Rathbone (1997) check. I have derived the indices by calculating the average price for each five year period using only prices which are reasonably certain and which have reasonably certain associated dates. A base price is selected and given the index value 1.00. The index values of all other prices are those prices divided by the base price.

was able to acquire wheat at the same price as for the period AD 60 to AD 215. Its ability to buy wheat at an unchanged price despite the debasement of the coins with which it paid for that wheat was lost by AD 250 at least and towards the end of the third century it paid just under 30 times as much per kilogram of wheat as it had done at the beginning of that period. That increase in pricing is broadly in line with the debasement of the currency.

Figure 7.2 State wheat prices



Figure 7.3 shows indexed Egyptian wheat prices. If changes in the silver content of the currency fully explain changes in price then observed wheat price indices, which are circled, should lie on the graph of the silver content index. There are, however, other factors at play. As outline above, the state intervened in the markets and bought as it required at prices which were fixed over long periods and which did not vary with changes in the annual supply of wheat or with changes in the competing demands for wheat.⁴⁰⁷ The presence of this major buyer with compulsive powers will have served to depress market prices. The extent of the

⁴⁰⁷ Through to AD 216 the state price for wheat was, with few exceptions, fixed at 8 drachmas the artaba.

depression will have varied with the scale of the market intervention.⁴⁰⁸ We also expect random year on year variations in the price of wheat because of random weather shocks. Such variations might be up to 30% of the underlying mean. The indexed wheat prices shown for the period up to the end of the second century, during which time the silver content of the denarius stayed broadly unchanged, are stable with variations that may be weather related. Towards the end of the second century and into the early third century there is a step change in price levels which broadly matches the level of debasement.⁴⁰⁹ Prices in the following half century remain broadly at this level despite continuing debasement of the coinage and become highly variable. Possible reasons are either a substantial increase in the supply of wheat or a reduction in demand. It is also possible that price levels are being influenced by the silver content of older coinages than just that of the newly minted coins. The increase in prices implied by the last price observations at the end of the period being examined broadly follows the trend implied by the debasement theory. The level of the increase, one order of magnitude, is such that this observation although related to very small numbers of transactions in a period of years is statistically credible.

⁴⁰⁸ If I am a buyer willing to pay 12 but knowing that if you do not sell your stock to me it will be bought by the state at 8 then I will offer something much less than 12.

⁴⁰⁹ Rathbone and van Reden (2015) 178 note the broad scholarly agreement that the doubling of Egyptian wheat prices between the AD 160s and the AD 190s is attributable to the Antonine plague but also note the uncertainties. The plague occurred in the period AD 165 to AD 180.

Figure 7.3 Private wheat prices

Figure 7.4 below shows the variations in wine prices in Roman Egypt. As with wheat the price of wine is dependent on the weather which will affect the size and quality of the crop. Unlike the case of grain, however, there can be considerable variation in the quality of the individual product. Price variability is therefore higher than for grain. Wine prices again show a long period of price stability up to the end of the second century. Price data from the middle to the end of the third century shows an increase in the level of prices but not to the extent suggested by the debasement theory.

Figure 7.4 Private wine prices

Figure 7.5 below shows changes in the prices of donkeys. These animals were used in the transport of goods and commodities, such as grain, and their price has some of the characteristics of capital-good prices. The supply of donkeys to the market is not dependent on the weather and is more likely to be affected by actual or anticipated economic conditions. Unlike grain, which is generally uniform in quality, the characteristics of the commodity being exchanged in individual transactions are variable. This introduces randomness into prices which we would expect to be greater than the weather related variation seen in wheat prices.⁴¹⁰ The price depression caused by state intervention is very unlikely to have been a major factor here. Again there is broad price stability until about AD 200 when there is again a phase shift in prices. The increase seen in prices in the period AD 200 to AD 230 is

⁴¹⁰ For one of these animals to be twice the price of another is unremarkable. Weather conditions need to be extreme for the price of grain to double.

somewhat above that implied by the debasement theory. The last two observations are strikingly similar to the corresponding wheat observation.

Figure 7.5 Price of donkeys

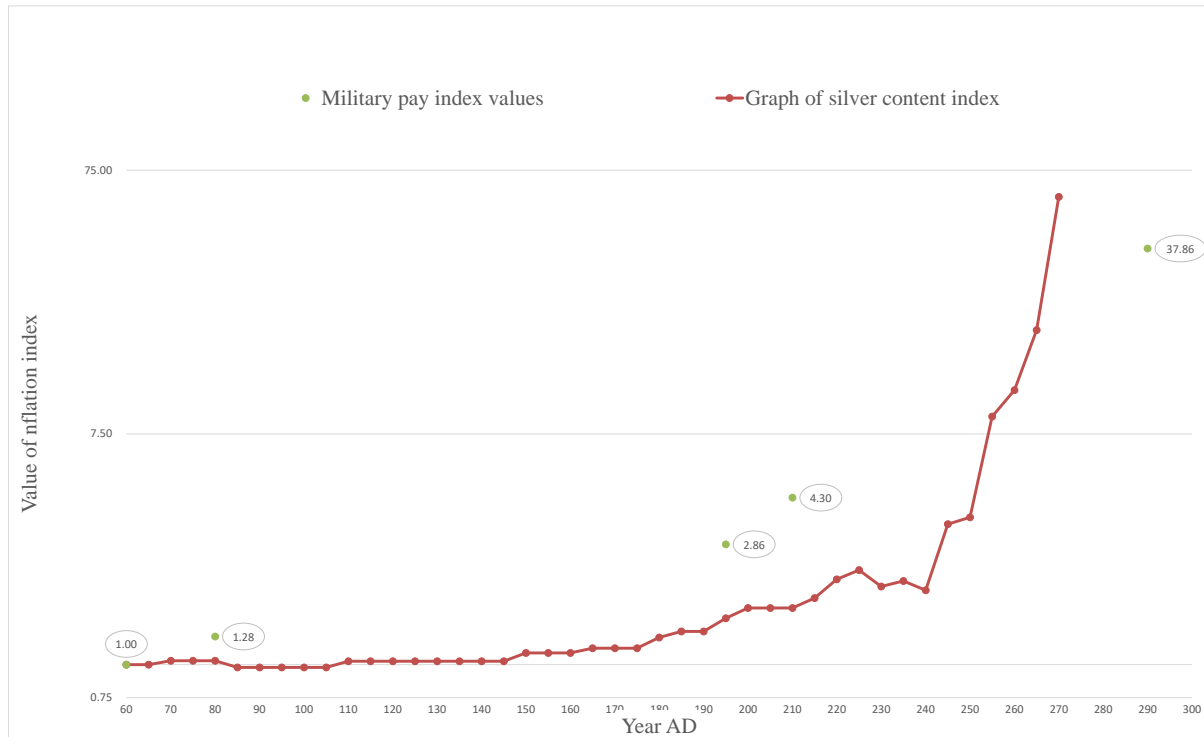


Finally figure 7.6 shows the trend in legionaries' pay.⁴¹¹ This wage inflation index does not show the random year on year variations already noted. The wages are set according to scales which are seldom changed. However, from the beginning of the third century a greater part of the remuneration of the soldiers was additional to the fixed scale and increase to the higher figure shown for the end of the third century took place gradually. The general features are as already observed. A long period of price stability, followed by an inflationary period and then by a period of hyperinflation. Around the middle of the third century severe erosion of military pay started to take place. The increases prior to that were substantially higher than the increases in the prices of wheat, wine and donkeys implying a substantial increase in the

⁴¹¹ The derivation of these figures is set out in Chapter 8. The military pay considered here is fixed pay plus additional benefits including donatives..

transfer of resources to the army. This advantage had been lost by the end of the third century AD.

Figure 7.6 Military pay



In general, it appears that until the end of the second century there was limited debasement of the currency and limited price inflation in items such as wheat, wine, donkeys or military pay. Starting from the end of the second century AD there is an emerging price instability and the level of prices broadly doubles over the period to AD 270. Around AD 270 to AD 280 prices increase approximately tenfold.

Price changes do not mirror exactly the debasement of the coinage but the general trends in price and in currency debasement are similar. The debasement of the currency is a relatively poor short-term explanatory theory since there are other factors at play. Nevertheless the long period of price stability in the first two centuries of the Principate is associated with a currency which suffers little debasement and it is likely that from the end of the second century the silver content of the currency no longer acted as a price constraint.

7.7 Conclusions

The monetary system functioned well through the first two centuries AD. The operation of the mints, which acted as instruments of political control, was consolidated and rival currencies were largely eliminated. Through this period there was no inflation of any significance and coins of this money system were recognised and used throughout a large area.

The inherent instability of a metallic currency did not manifest itself through the period of the first two centuries AD. The fixed relationship between the aureus and the denarius remained current until about the middle of the third century AD, although there had always been pressure on the relationship between the as and the denarius. This further aspect of currency stability very likely followed from the centralisation and control of the mint and the suppression of rival currencies.

The functioning of large parts of the economy depended on the availability of currency. Controllers of production on agricultural units which were constructed to produce purely for external consumption followed strategies which required a monetary system. The output of these estates was sold for money largely into the urban centres which were of a size that required a money based system for their inhabitants to function economically. The trade that was carried out over the large expanse of the empire also required a monetary system and much of it is very likely to have required a banking system. So long as the monetary system of the empire was intact these urban and trading structures could continue to operate and continue lead to produce and exchange of goods through the empire. The stability of the currency greatly facilitated the approximately fivefold increase in trade which occurred in that period.

This long period of monetary stability and wide acceptance of a single currency is probably unique.

CHAPTER 8: The system of state finances and fiscal incompetence

8.1 Introduction

This chapter is about the mechanism by which the state funded itself. The purpose of this chapter is to illustrate the funding mechanisms which supported the state's contribution to the economy as illustrated in earlier chapters. These contributions included the provision of infrastructure, such as ports and canals, which supported trade and aqueducts which supported urbanisation. The army which defended the empire, maintained security within its borders, provided manpower for infrastructure development and was a mechanism through which knowledge was deployed depended on the state for its finance. Impairments in the state finances had consequences for the economy.

This chapter explains the dynamics of the funding structure, the institutions which underlay that structure and the weaknesses in the funding structure. The resilience of the system, its capacity to continue to function as circumstances changed, is examined. The financing structure was sophisticated and depended on individuals who were schooled in administrative, financial, legal and other matters. In this chapter I suggest that the weakening in the process by which this expertise was embedded in individuals allowed catastrophic financial decisions to be made which broke the central control of the state.

To bring clarity to the analysis I emphasise the distinction between current and capital expenditure. Generally speaking, increases in current expenditure have continuing effects since, for example, it is difficult to remove a pay increase once it has been granted.⁴¹² Capital expenditure is discretionary and can go up – as under Claudius – or down – as under Tiberius.⁴¹³ Equally, in this chapter I work to ensure that the different types of expenditure are given their proper perspectives. Lavish state expenditure on infrastructure projects, although more visible and impressive in its effects, can be of less financial significance than small increases in salaries to state dependents.

The functions of the state which are supported by this funding structure are briefly described. These functions depended on a balance between the income available to the state and the expenditure of the state. While through most of the first and second centuries AD the regular income from the tribute of the provinces had broadly matched current expenditure on such

⁴¹² Exceptions to this point are one-off current expenditure on handouts to the population, for example.

⁴¹³ Although, some capital expenditure represents long term commitment, as for instance, in the construction of ports.

items as soldiers' pay and provincial administration the balance between total income and total expenditure had from time to time been disrupted by occasional excessive expenditure on non-recurring items. The underlying finances were, however, robust. This balance began to be disrupted at the end of the second century and was destroyed through a series of executive decisions – which need not have been taken but were irreversible – in the early part of the third century. The destruction of the balance between current income and expenditure did not lead to a collapse of the state but broke the central control which was never fully restored. A partial solution emerged under Diocletian who, in the construction of the tetrarchy effectively split the empire into two regions in one of which the numbers still worked and in which central control was still possible and into another region where the numbers no longer worked and where different state structures had to be constructed.

8.2 The debate

The debate around the finances of the state centres on elements of those finances rather than on attempts to understand the balance between overall income and expenditure and how this balance changed over time. As will be seen one of the largest items of state expenditure was the cost of the army and much detailed work has been done on the pay rates and overall cost of the army. There is continuing debate on the rate at which pay increased under Septimius Severus and debate as to the pay relativities of legionary soldiers and auxiliary soldiers.⁴¹⁴ Some work has been done on the income of the state but this is usually of a very cursory nature, reflecting the limited data available.⁴¹⁵

8.3 Funding structures and flows

Two main operations provided funding for the state. The first of these was the *aerarium saturni*, which was a state function, and the second was the *patrimonium*, which was the property of the Emperor. There were others such as the *aerarium militare* which are described later. The *patrimonium* consisted of the estates of the emperor and was initially held as his personal property to be disposed of by inheritance as he wished. In time this changed and the *patrimonium* was held by the emperor as by right of office. It was essentially an asset base consisting of estates that he held and could be added to by conquest, by receipt

⁴¹⁴ See Frank (1940), Develin (1971), Duncan-Jones (1978), Hopkins (1980), MacMullen (1984) and Rathbone (2007) for discussion of army pay. See M.P. Speidel (1973), M.A. Speidel (1992) and Alston (1994) for discussion of these points as well as commentary on army pay.

⁴¹⁵ See Frank (1940), Hopkins (1980), Duncan-Jones (1994) and Rathbone (1996b).

of legacies and by confiscation. There were other sources of income to the *patrimonium* such as the estates of those who died intestate.

The Roman state did not tax its citizens directly and land held in mainland Italy, for example, was tax exempt. The collection of direct tribute taxes from non-citizens was organised through the cities and tribes of the empire, from which an amount of tax was collected calculated from tax rates applied to the tax base associated with the entity. In addition to the direct tribute tax there were various indirect taxes, the collection of which was generally farmed out to *publicani* in the early empire as it had been in the Republic. These indirect taxes were sometimes hypothecated to particular state functions or liabilities. The 5% inheritance tax and the 1% auction tax went to the *aerarium militare* for the funding of the soldiers' bounty. The 4% tax on the sale of slaves was used to pay for the city police. There were other taxes which were not so allocated. These included the port dues paid on imports and exports which were 25% on luxury goods at the Red Sea and elsewhere varied between 2% and 5% and the manumission tax of 5%.⁴¹⁶

Not all taxes were determined as money amounts and much was collected in kind, although the amounts collected in kind varied by province. The collection of tax in kind was advantageous to the extent that part of the liabilities of the provincial authorities were determined in kind. For example, a substantial part of the cost of the locally based army units was required for grain and some of the liabilities of the central authorities at Rome were for the meeting of their grain needs there. To the extent that the tax income streams matched these liabilities the arrangement was advantageous although there was not always a precise match.⁴¹⁷ The essential difference between the collection of the tax in money and in kind was that collection in money required the taxpayers or the responsible local elite to sell produce for coin to buyers – who might include the provincial authorities – whereas the collection of at least some of the tax in kind reduced some of the supply risk of those authorities.

8.3.1 The continuing revenues and their importance

Various attempts have been made to quantify the amounts of tax revenues collected. In 1920 Frank estimated that the state revenues had been about HS 500 million in the middle of the reign of Augustus and secondly estimated that, based on a survey of literary sources, the

⁴¹⁶ See Rathbone (1996b) 312-6 for a discussion of the tax system.

⁴¹⁷ For example, if the tax due was expressed as a percentage of crop yield surplus or loss arose as the yield rose or fell.

imperial and senatorial revenues had been between HS 1,200 million and HS 1,500 million by the time of Vespasian.⁴¹⁸ In 1980 Hopkins gave his opinion that the first of these two estimates was too low and that the second estimate of HS 1,200 million to HS 1,500 million was closer to the truth.⁴¹⁹ Duncan-Jones in 1994 suggested a figure for tax revenues of about HS 800 in the AD 70's.⁴²⁰ Rathbone in 2007 gave an opinion that this estimate was about 50% too low because it omits the income from indirect taxes.⁴²¹ Estimates of likely tax revenues need to be time specific and need to consider the conditions in which the tax is collected. In favourable political and economic times tax revenues will tend to increase and conversely. The tribute that could be collected from the Empire also varied as the size of the Empire varied. In order to get a broad fix on the amount of tax that might have been available I will consider the tribute, the indirect taxes and the income that came from the *patrimonium* separately because it is easier to estimate at these separate levels and because identifying the relative importance of these components will make discussion about the resilience of the system clearer. I estimate the revenues from the tribute as

$$\text{Tribute Revenue} = \text{Number of Provinces} \times \text{Average Tribute} \quad \text{Formula 8.1}$$

The first element of this equation is easy to establish and the second element can be selected based on a review of the literary sources which contribute by either giving precise amounts that were collected from a variety of provinces or by giving statements as to the relative contribution different provinces made. Plutarch gives figures for the revenue of the state before and after the conquests of Pompey, in the middle of the first century B.C.

In addition to all this, the inscriptions set forth that, whereas the public revenues had been HS 200 million, they were receiving from the additions which Pompey had made to the city's power HS 340, and that he was bringing into the public treasury in coined money and vessels of gold and silver HS 480 million, apart from the money which had been given to his soldiers, of whom the one whose share had been the smallest had received HS 6000.⁴²²

These figures indicate a dramatic increase in public revenues, which are indicated to have more than doubled from HS 200 million to HS 540 million. To place this in context, it would

⁴¹⁸ Frank (1920) v.53.

⁴¹⁹ Hopkins (1980) 116-7.

⁴²⁰ Duncan-Jones (1994) 46.

⁴²¹ Rathbone (2007a). 175.

⁴²² Plutarch, *Pompey* 45.3. I have translated monetary amounts at the rate of 1 drachma is HS 4 and 1 talent is HS 24,000.

have been impossible to have funded the army under the pay structure established by Caesar from the recurring revenues of the state before Pompey's conquests.⁴²³ The revenue available after the conquests of Pompey could fund such a pay structure and leave significant surplus. The revenue which Plutarch indicates was made available to the state from booty, which is one-off revenue, was HS 480 million or less than one year's recurring income. The value of the provinces to the state was not so much in the one-off revenues that could be extracted on conquest but in the continuing revenues. The bounty made available to the individual soldiers was at least HS 6,000. In the discharge scheme implemented by Augustus for soldiers on retirement after 20, later 25, years' service and described later in this chapter, the discharge bounty was HS 15,000. The bounty that Pompey made available to his soldiers was just about seven years' pay at the rate implemented by Caesar. The figures given by Plutarch indicate a revenue per province of between HS 40 million and HS 50 million. The figure given by Suetonius for the revenue under Caesar are consistent with what Plutarch has to say about revenues under Pompey.

During the nine years of his command, this is the substance of what he did. All that part of Gaul which is bounded by the Pyrenees, the alps and the Cevennes, and of the Rhine and the Rhone rivers, a circuit of some 3,200 miles, with the exception of some allied states, he reduced to the form of a province and imposed on it a yearly tribute of 40 million sesterces.⁴²⁴

The figure given by Josephus for the province of Judaea in the time of Claudius (AD 41 to AD 54) is similar to these first two figures but is very likely exaggerated since it applies to a province which was incapable of producing the revenue of Gaul.

The revenues that he received out of them was very great, no less than HS 48 million. Yet did he borrow great sums from others, for his expenses exceeded his income and his generosity was boundless.⁴²⁵

Philostratus gives a figure for the time of Hadrian (AD 117 to AD 138) which is similar to but lower than the figures of Plutarch and Suetonius.⁴²⁶

⁴²³ See Table 8.6 later.

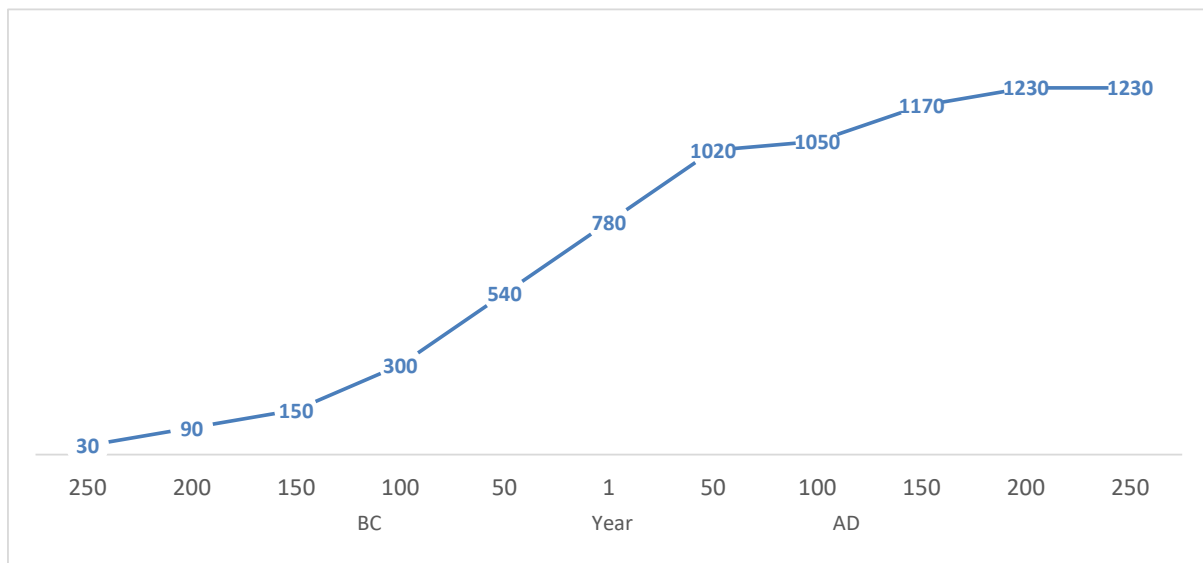
⁴²⁴ Suetonius, *Julius Caesar* 25.1 indicates that the revenue of Gaul in 49 B.C. was HS 40 million.

⁴²⁵ Josephus, *Jewish Antiquities* 19.352 indicates that the revenue of Judaea under Claudius was HS 48 million.

⁴²⁶ Philostratus, *Lives of Sophists* 548 indicates that the revenue of Asia under Hadrian was HS 28 million.

On the basis of these sources I select HS 40 million as the amount that on average was acquired as ongoing tribute revenue when a province was established from conquered territory. The following graph shows the development of tribute income as the territory under the control of Rome expanded. The income shown in the time of Pompey is between HS 500 million and HS 600 million which happens to agree with a literary source.⁴²⁷ The tribute income for the middle of the reign of Augustus is about HS 800 million and by the end of the second century AD has increased to over HS 1,200 million.

Figure 8.1: Schematic of the current account revenues from Tribute over time



Of the two components, number of provinces and average revenue per province, the first is the more certain. The shape of the graph, which depends on the number of provinces, is also certain and illustrates a change from a conquest model. The graph is steeply increasing in the period until the middle of the first century AD. Through this period there are substantial one-off benefits to accrue to the state and to those involved in conquest. These benefits diminish in the first century AD which becomes a transition period from a conquest model to the trade model already described. I believe that the implementation of the pay structures under Caesar may have been part of the process of managing this transition. Stipendia were introduced as the bonus flows from conquest reduced. This transition had more to be managed at the level of the soldier rather than at the level of the public finances. By the first century AD the benefits from booty to the state were substantially below those available from the continuing

revenues.⁴²⁸ Investing in port infrastructure becomes more remunerative than conquering provinces. Protecting the existing revenues of the state was more important than attempting to add to revenues through conquest.

I find the estimation of the indirect income to be much more difficult because it relies on a quantification of the volumes of trade. To an extent the task is made easier by the fact that we do not have to estimate the income that came from the tax headings that were allocated to specific liabilities; we can treat these as separate revenue flows. We are then left with the port dues and the income from the mines. I believe that trade had increased substantially and probably about twenty times over the period from the beginning of the Principate to the start of the third century AD. At a guess trade revenues were probably about HS 5 billion and the tax on this was probably between HS 250 million and HS 500 million.

These figures suggest that the recurring tax revenues of the state during the Principate were about HS 1,000 million to HS 1,500 million with the latter figure applying as trade increased in the second century AD. These figures are not far distant from the more recent scholarly estimates.⁴²⁹

It is possible to very broadly estimate the income that might have been produced from the *patrimonium* by consideration of the expenditure detailed in the *Res Gestae* of Augustus. The *patrimonium* is different in its potential contribution to state expenditure in that it owned physical assets rather than simply the right to receive income. It could contribute to state expenditure by the sale of parts of the asset base as well as by the deployment of income generated from the asset base. The expenditure that Augustus lists as having been his personal contribution to the state extends from the erection of buildings to contributions to the army. Since his purpose is to make clear the extent and magnificence of his achievements we do not expect much to have been omitted. Augustus' total personal expenditure amounts to HS 2,400 million over his reign.⁴³⁰ This is about HS 60 million per annum which I will assume is entirely made from income rather than asset sale. This is equivalent to the average tribute revenue of something under two provinces. The asset base that produced this annual income was between HS 1,000 million and HS 2,000 million in value.⁴³¹ While this asset

⁴²⁸ If each conquest brings in booty equivalent to the revenue of two provinces then an additional province, when you have thirty provinces, brings in booty which is equivalent to less than 10% of recurring revenue and available for one year only.

⁴²⁹ The conclusions of this chapter are not sensitive to this assumption.

⁴³⁰ See Frank (1940) v.14-6 for a review of the figures published by Augustus in the *Res Gestae*.

⁴³¹ Assuming a yield of between 3% and 6%.

base could significantly contribute to shortfalls in state income or temporary rises in expenditure through asset sales the regular contributions that the *patrimonium* could make to the support of the state was limited and substantially less than came from the tribute and from indirect taxation.

8.3.2 Other revenues

Other revenues which were outside the normal tax collecting systems flowed to the benefit of the state. These revenues include gains from the sequestrations which could happen in the normal course of events where, for example, the assets of a man who committed treason were seized. The value that could accrue through sequestration was potentially considerable. However, to monetise this was not a trivial matter and arbitrary or extensive confiscation of assets was likely to be destabilising.⁴³² The confiscation of an estate valued at HS 300 million added about 20% to the revenue of the state but only for the year in which it occurred. Unless expenditure dropped then further similar confiscation was required.

Conquest was another source of revenue although in the period of the Principate and certainly after the first century AD only limited territory was added to the Empire and this source of revenue, which unlike tribute was non-recurring, was limited. The state was therefore essentially funded from tribute and indirect taxes. By the time of the Principate all other sources of revenue were insignificant in comparison to those recurring revenues. The stability of the state finances depended on the balance between the ongoing revenues and the expenditure of the state.

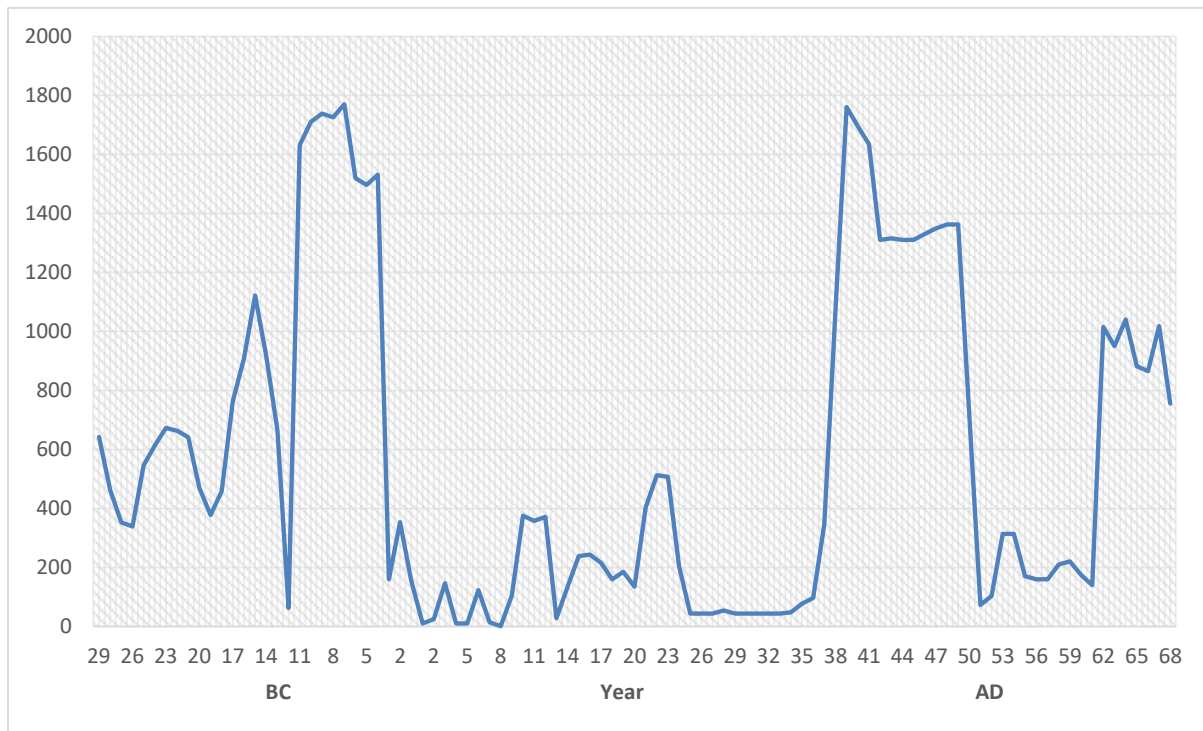
8.4 Capital expenditure

M.K. and R.L. Thornton in 1989 estimated the total number of work units spent on capital projects under the Emperors Augustus through Nero.⁴³³ These work units are intended to provide a measure of the relative amount of work undertaken over time. The results of their work are summarised in the following graph which illustrates the extent to which capital expenditure varied over time. The units of measurement used are the Thornton work units.

⁴³² Where the estate consisted of coin or bullion or objects made out of precious metal the process of monetisation was trivial. When the estate consisted of landed property then if these were dispersed then multiple sales to different owners could be affected. The transfer of an estate of HS 300 million from one owner to another in exchange for comparable cash was always going to be difficult. The transfer took place in the context of confiscation and the recipient knew when he received this asset and parted with his cash that he was a tempting target. The difficulty of disposing of large blocks of assets is attested by the difficulties faced in the liquidation of the estates of Augustus to pay for the legacies he had left in his will.

⁴³³ In this subsection I rely heavily on M. Thornton and R. Thornton (1989).

Figure 8.2 Illustration of variations in capital expenditure over time – in work units



The nature of the construction projects being undertaken varied over time. For example, the work undertaken by Augustus related to the construction of places of public assembly - theatres and forums - and also to the repair of old aqueducts and the construction of new aqueducts. Much of this work was undertaken by the same workforce moving to new projects as old projects were completed and largely was completed by 4 B.C. Capital expenditure under Tiberius was minimal but there was a modest increase under Caligula who initiated work on new aqueducts, which were completed under Claudius, and who undertook some work on palaces.

There was a significant increase in capital expenditure under Claudius, whose reign started in AD 41. Much of this work was infrastructure related and significant projects included the draining of the Fucine lake and the construction of the port at Ostia as well as the construction of aqueducts. Claudius built a tunnel of 5.6 kilometres in length to create about 15,000 hectares of fertile land. This work was exceptional only for its size and cost and not for any innovation. The project took 11 years and employed about 3,000 direct labourers. The construction of the harbour of Ostia, which started under Claudius, was intended to facilitate the grain supply to Rome and to reduce famine risk. The project lasted from AD 46 to AD 64 with the completion of warehouses and other harbour buildings happening under Nero. The work undertaken by Nero was more focussed on entertainment, rather than economic

infrastructure, and related to the construction of baths, amphitheatres, the rebuilding of the Circus Maximus and the reconstruction that followed the fire at Rome.

In order to place these different work streams, which were focussed on different purposes, into context I have converted the Thornton work unit estimates into monetary amounts.⁴³⁴ The following table summarises the results of these calculations. The amounts are in HS millions.

Table 8.1 Summary of Capital Expenditure

1: Emperor	2: Reign End	3: Total Expenditure	4: Average Annual Expenditure
Augustus	AD 14	1,300	30
Tiberius	AD 37	190	8
Caligula	AD 41	300	77
Claudius	AD 54	600	46
Nero	AD 68	390	28

As can be seen the average annual expenditure through the period is highly variable and reflects the approaches taken by the different emperors. More significantly the amounts are very low relative to the total recurring revenues of the state of about HS 1,500 million and annual capital expenditure varied between 0.5% and 5.0% of recurring tax revenues.⁴³⁵ The revenues consumed were less than the tribute of one province. These significant and impressive projects, the port at Ostia and the forum of Augustus, for example, were not

⁴³⁴ They estimate the cost of the two Claudian aqueducts at 7000 work units each. The **literary sources** indicate that the Claudian aqueducts cost HS 350 million each and so I equate each of their work units to HS 50,000 equivalent. This conversion effectively assumes that the relationship between labour and material costs is broadly the same across all the construction projects, which in aggregate I think is a reasonable assumption to make. Note that Pliny *Ep* 37 gives the expenditure on the aqueduct started by the citizens of Nicomedia but then abandoned as HS 3,318,000, which he considers an enormous sum. Presumably this was much smaller than any of the Roman aqueducts.

⁴³⁵ Current US figures are about 10%.

difficult to finance. In part this is because the manpower required was low relative to what was needed for the army. As already noted, the draining of the Fucine lake probably needed about 3000 men for a number of years. This is just over half the strength of one legion and there were 30 legions and as many auxiliaries again in the army. Serial construction meant that a labour force of that size could produce substantial infrastructure over time at a cost which was minimal relative to the cost of the army.

The evidence from building inscriptions, not all of which relate to state expenditure, shows that construction continued during the second century AD and at a higher level than seen in the first century AD. Construction then declined dramatically in about 240 AD. Honorific inscriptions also show a decline around the same time.⁴³⁶

These projects had something of an effect in boosting economic performance through employment although this was limited both because of the limited number of men involved in these projects and because in some cases the work was undertaken by soldiers and therefore did not generate additional employment. The economy benefited more in the opportunities for trade, for example, that were possible because of additional port infrastructure. The impressive monuments of Rome were not difficult to finance.

8.5 Current Expenditure

The main items of recurring expenditure were the costs of the army, of provincial government and of the grain dole. Nonrecurring items of current expenditure included the one-off provision of lavish games, for example, or the disbursement of sums of money to the city population. These one-off items of expenditure could have significant effects. For example, a gift of HS 1,000 to each of those receiving the grain dole together with the Pretorian Guard could cost HS 200 million and change a budget surplus into a budget deficit.⁴³⁷

Nevertheless, one of the major items of recurring revenue was the cost of the army which because of the extensive scholarly work that has been undertaken on manpower and on salaries can be estimated to a reasonable degree of accuracy. Its estimation also facilitates a deep understanding of the dynamics of state finances.

⁴³⁶ Wilson (2011) 163-7 based on data from the *Epigraphische Datenbank Heidelberg*.

⁴³⁷ Dio 77.1.1 costs the donation that Septimius Severus gave, on the tenth anniversary of his accession, to all those in receipt of the grain dole and to the Praetorian Guard, of one gold piece for each year of his reign at HS 200 million.

8.5.1 Army manpower

The army consisted of legions, in theory recruited from Roman citizens, and auxiliary forces, recruited from provincial subjects. The number of legions varied somewhat through the Principate. There were twenty five legions in the middle of Augustus' reign and either twenty eight or twenty nine by AD 70.⁴³⁸ By the time of Diocletian the number of legions was about thirty. Legions were about 5,000 men strong but legion size could be flexed through increased or reduced recruitment or through changes to the mandatory period of service. The estimation of the number of auxiliaries is somewhat more problematic. Tacitus implies that they numbered about the same as the legionaries in the time of Tiberius. It is likely that over time the auxiliaries became a larger part of the force and Holder gives an absolute figure of 250,000 which I will use as the figure for late Severan times.⁴³⁹ The following table sets out the numbers of soldiers employed by the army and how this varied over time.

Table 8.2 Variations in army manpower over time

	Augustus	Domitian	Septimius Severus	Caracalla	Diocletian
	27 B.C. –AD 14	AD 81 -96	AD 193- 211	AD 198-217	AD 284-305
Legions	25	28	30	33	33
Legionaries	125,000	140,000	150,000	165,000	165,000
Auxiliaries	125,000	140,000	150,000	250,000	250,000
Total Soldiers	250,000	280,000	300,000	415,000	415,000

8.5.2 Army pay accounts

A small number of papyri survive which document soldiers' accounts and the deductions made from their pay.⁴⁴⁰ These accounts follow two different forms and provide substantial insights into the funding costs of the army. In the first accounting format deductions were

⁴³⁸ Keppie (2005) 388.

⁴³⁹ Holden (1980).

⁴⁴⁰ Woolf (1999) 53 notes that soldiers who learnt to read might apply that to commerce. Equally soldiers had the opportunity to learn, from an examination of their pay accounts, systems for recording credits and debits, reckoning balances and carrying amounts forward all of which is helpful to successful commercial ventures.

made for the costs of hay, food and, from time to time for clothing, for example.⁴⁴¹ The following table shows the pay accounts, which follow the first format, for the auxiliary soldier Q. Iulius Proculus who was born in Damascus and is dated to AD 81.⁴⁴² The amounts shown are in HS. The year is divided into three accounting periods.

Table 8.3 Example of soldiers pay accounts

	Accounting Period 1	Accounting Period 2	Accounting Period 3	Annual
Salary	247.5	247.5	247.5	742.5
Hay for bedding	10	10	10	30
Food	80	80	80	240
Boot, socks	12	12	12	36
Camp Saturnalia	20	0		
Clothes and tailoring	60	0	145.5	205.5
To the standard	0	4	0	4
Total Spent	182	106	247.5	535.5
Deposited	65.5	141.5	0	207
Opening balance	136	201.5	343	136
Closing balance	201.5	343	343	343

⁴⁴¹ *RMR* 68;*RMR* 69;*RMR* 70.

⁴⁴² *RMR* 68.

The papyrus also records the salary deposited for his colleague C. Valerius Germanus who was born in Tyre. The deposit is the same but for the second soldier the clothing deductions total HS 245.5. This second soldier also had a smaller opening balance and closing balance.

Accounts which survive from AD 84 follow a similar format but are less well preserved.⁴⁴³ These accounts date to after the Domitian pay increase, which is described later. What is noteworthy is that the deductions have also increased broadly in line with the increases in salary. In effect by increasing these deductions an apparent 33% increase in pay was in reality an affordable 10% increase.

The accounts that survive from the end of the second century and from the beginning of the third century show a different and simpler format. They itemise two fund amounts with the camp authorities (a deposit and a viaticum), they indicate the amount received as from the stipendium, they show a deduction of about 6% for tax and then they show the balance left over. In some cases balances are negative. The pay deposited has increased significantly.⁴⁴⁴ No deductions are shown in these accounts for hay, food and so forth. The change in the type of accounts being kept moves, therefore, from a form which shows substantial deductions to one which shows no deductions for anything other than tax. This change indicates that at some stage before AD 192 (which is the earliest date associated with the new format) the pay structure of the army had changed so that soldiers no longer had the cost of food and other items taken from their salaries. This change represented a substantial increase in soldiers' remuneration.⁴⁴⁵ Equally, future pay increases could not be offset by increases in deductions and the amount of cash flowing to soldiers increased enormously.⁴⁴⁶

The first accounting format illustrates some of the funding structure of the army. Taking the monetary values at face value then about a third of the salary cost is funded by the provision of basic requirements such as food and a further third was funded through the provision of clothing. The provision of these items essentially represented two subsidiary businesses of the army and the output from these businesses covered about two thirds of the cost of the army through the first and second centuries AD. It is very likely that the amounts being

⁴⁴³ *RMR* 69. The stipendium shown is equivalent to HS 297 or to an annual deposit of HS 891.

⁴⁴⁴ *RMR* 70/*ChLA* 410 dated to AD 192 show deposits of HS 228.25 equivalent – which correspond to annual figure of HS 1,014.75 – and *ChLA* 446 and 495 which are dated to the second or third century shows deposits of HS 1031.75 equivalent – which corresponds to an annual figure of HS 3,093.75.

⁴⁴⁵ I, unlike other authors, allow for this effect in calculating the cost of the army. I conservatively value this at 1/3 of base pay.

⁴⁴⁶ Removing the deductions and doubling pay, for example, means that soldiers' cash increases about sevenfold.

deducted are notional charges. They are round numbers and do not vary through the year. Profit or loss therefore arose depending on the extent to which the cost of acquiring these items differed from the notional charges recouped from the soldiers.⁴⁴⁷ To the extent that these goods were received as taxes in kind then the profit or loss arose by reference to what they could alternatively have been sold for in the market.

8.5.3 Pay rates

The pay structure established by Caesar consisted of a basic pay rate, which was paid in three instalments each year and a set of pay relativities which determined how much some higher paid soldiers got.⁴⁴⁸ This simple structure of a basic pay rate and a set of relativities to it, with payments made three times a year, survived at least into the time of Diocletian.⁴⁴⁹ The basic rates of pay that prevailed from the time of Caesar until the time of Septimius Severus are well established from the literary sources. The evidence from Tacitus indicates an annual pay rate of HS 900 equivalent under Caesar and is consistent with the evidence from Dio when he describes the increase in pay from the equivalent of HS 900 to the equivalent of HS 1,200 that took place under Domitian in AD 84.⁴⁵⁰ This rate of increase of one third is also consistent with the Suetonius description of Domitian increasing the pay from three instalments of HS 300 equivalent to four instalments.⁴⁵¹ The next pay increase took place under Septimius Severus in AD 197. All that the one source which refers to this increase tells us is that the increase was the greatest yet seen; neither the amount nor the rate of increase is quantified.⁴⁵² It has been variously argued that pay increased to amounts between HS 1,600 per annum and HS 2,400 per annum.⁴⁵³ The difference in these figures is to an extent a difference in timing rather than quantum, since it is clear that pay eventually exceeded all these figures.⁴⁵⁴

⁴⁴⁷ By the end of the third century this mismatching profit or loss no longer existed. Much of his basic food requirements were provided to the soldier directly as payments in kind.

⁴⁴⁸ It is possible that the introduction of the pay based army remuneration structure led to increased monetisation.

⁴⁴⁹ See Duncan-Jones 1978 for a discussion of pay under Diocletian.

⁴⁵⁰ So, for example, mutinying soldiers are described by Tacitus as complaining about their pay rates of 10 asses a day. Tacitus *Ann* 1.17.4. Dio 67.3.5.

⁴⁵¹ Suetonius, *Dom* 7.3.

⁴⁵² HA, Sev 12.2; Herodian 3.8.5.

⁴⁵³ Develin (1971) and Duncan-Jones (1990) estimates give HS 1,600. Watson (1959) and Alston (2007) estimates give HS 1,800. Brunt (1950) and Speidel (1992) estimates give HS 2,000 and HS 2,400 respectively.

⁴⁵⁴ If we follow Duncan-Jones (1978) then we know that the size of the stipendium at the end of the third century is HS 4,800. So assuming smaller increases under Septimius Severus means we must assume larger increases after him.

I will take the view that Septimius Severus doubled the pay rate following the evidence from an inscription from Lambaesis. This inscription contains a dedication to Septimius Severus set up by a *cornicularius* and an *actarius* from a detachment that served in the Parthian campaign grateful for the doubling of pay that he had instituted for the *legio III Augusta*.⁴⁵⁵

The long period of pay stability that preceded this increase was over and very shortly thereafter there was another substantial increase when Caracalla added 50% to the basic pay rate in AD 212.⁴⁵⁶ By the time of Diocletian the basic pay rate had been increased to HS 4,800 although the precise date of this third century increase is unknown.⁴⁵⁷

The following table summarises my assumptions as to the basic annual pay rate of legionary soldiers in HS. It is clear that from the end of second century the burden that the financing of the army placed on the continuing revenues of the state is being dramatically increased.⁴⁵⁸

Table 8.4 Basic pay rates that applied under various Roman Emperors

	Augustus	Domitian	Septimius Severus	Caracalla	Diocletian
Reign	27 B.C. –AD 14	AD 81 -96	AD 193- 211	AD 198-217	AD 284-305
Basic annual pay rate	900	1,200	2,400	3,600	4,800
Rate of increase		33%	100%	50%	33%

⁴⁵⁵ Rathbone (2007a) 160. Dessau (1908) 462-3.

⁴⁵⁶ Herodian 4.4.7. Macrinus complains that the subsequent increase under Caracalla cost HS 280 million. Dio 78.36.3. Maximus Thrace may have doubled the basic pay rate in AD 235.

⁴⁵⁷ Duncan Jones reconstructed the pay rate that applied by the time of Diocletian from the Beatty papyri of Panopolis, two of which papyri give the total amounts of wheat and barley supplied in A.D 228 as a two-month ration to a cavalry unit based at Thmou, the total *annona* paid to that cavalry unit for a four-month period in AD 299 and the total stipendium paid for a four-month period in AD 300. With an assumption as to how much the ration of barley was for a horse the troop numbers can be reconstructed. This analysis yields a pay rate of HS 7,200 annual equivalent for a detachment of cavalry and cavalry were paid 50% more than legionaries. This gives a Diocletian basic pay rate of HS 4,800. See Duncan-Jones (1978) 542-5 and Rathbone (2007a) 162.

⁴⁵⁸ The reign of Caracalla and Septimius Severus overlap.

The amounts shown in the above table are the basic pay rates.⁴⁵⁹ Small numbers of soldiers in senior positions were paid more than these amounts and officers were paid substantially in excess of these amounts. The cavalry were paid more than foot soldiers, at least in part to allow for the costs of their horses. I will assume that these high pay rates added 25% to the total cost of the army.⁴⁶⁰

As already noted further significant changes to the pay structure also occurred, possibly in the time of Commodus, when the practice of making substantial deductions from soldiers pay stopped.⁴⁶¹ From then on soldiers were no longer had to 'buy' provisions from the state but were provided with them outside the cash salary system, which is a significant change in how remuneration is seen. By the time of Diocletian soldiers were also provided with additional *annona* payments. Substantial donatives were made and by the end of the third century a total of four donatives were regularly paid to the soldiers. Type A donatives were paid on the anniversary of the accession and of the birthday of the Augusti and amounted to HS 2,500 equivalent each and Type B donatives were paid on the anniversary of the accession and of the birthday of the Caesars and amounted to HS 1,200 equivalent each. The three stipendium instalments had become a very small part of the total cash compensation which mostly now came on the anniversary of the birth and on the anniversary of the accession of the Caesars and Augusti.⁴⁶²

⁴⁵⁹ There is some debate as to the relationship between the basic rate of pay for the legionary soldier and for auxiliary soldiers. The literary sources, when they make reference to the pay of soldiers, talk about legionaries and provide no evidence for the pay of the auxiliary troops. The pay of the auxiliary troops has been a matter of controversy with some scholars assuming that they were paid the same and others assuming that the pay of the auxiliary was five sixths of that of the legionary. The amount of annual pay deposit of HS 742.5 shown in *RMR* 68 which shows the accounts of an auxiliary differs from the HS 900 figure for pay. Speidel (1992) notes that this is 5/6 of 99% and assume auxiliary pay is therefore 5/6 of legionary pay. Alston (2007) however takes the view that the difference relates to the cash that the soldiers received. I will assume that the two categories were paid the same amounts, since I consider it difficult to pay different amounts to soldiers serving closely together.

⁴⁶⁰ As already noted not all soldiers however were paid the same. Some soldiers called *principales* were paid pay and a half or double pay but the numbers in these ranking were relatively small and the additional cost was also small, perhaps of the order of 5% of total salary costs. MacMullen (1984) 578 suggests these were about 7% of the legionaries. Rathbone (2007a) 161 n13 gives a figure of 2%. The pay of the officers that is of centurions, tribunes and prefects was much greater and is uncertain and although these officers were few in number the additional cost of these rankings was at most 25%. See Rathbone (2007a) 161 for a discussion of the relativities. The 25% figure is based on MacMullen's weightings and taking these as higher limits.

⁴⁶¹ *RMR* 70 which is discussed in the last subsection is dated to AD 192 and shows that deductions are no longer being made from soldiers' accounts. I associate changes in army pay in the Principate with political instability, see later, and therefore attribute this change to the reign of Commodus rather than earlier in the second century AD

⁴⁶² Duncan-Jones (1990) 105-17.

These additional amounts per soldier are an annona of HS 2,400 and donatives of HS 40,000, with in kind payment of grain, oil and salt by the time of Diocletian.⁴⁶³ The in-kind payments were about the annual needs of those items. The remuneration structure had now reverted to something that was more like what had prevailed before the conquests of Pompey and the fixed pay structure of Caesar. Cash pay was largely bounty associated with the leader, the amounts paid were not guaranteed, basic provisions were provided directly in-kind to the soldier rather than being charged for.

Since the time of Augustus legionary soldiers had received a discharge bounty which was initially set at HS 12,000.⁴⁶⁴ This was increased to HS 20,000 by Caracalla and replaced by donatives in the time of Diocletian or just discontinued.⁴⁶⁵ Augustus set up a separate treasury to pay these bounties with an initial endowment of HS 140 million and the ongoing cost was to be met by an inheritance tax of 5% on large inheritances outside close family and 1% auction tax.⁴⁶⁶ Discharges probably numbered about 6,000 in the time of Augustus.⁴⁶⁷

The pay increases shown for the first and second centuries took place in a period of very low inflation and represented real increases in the value being captured by the army. Septimius Severus had seized power in AD 193, a year in which there had already seen two emperors come and go. The increase in pay under Septimius Severus may be happening because there is a need for the emperor to control the army and this is a way for him to do it or the giving of a pay increase might have led to political instability.

8.6 State accounts

The following table sets out the cost of the army under various Roman Emperors. The total army costs include the cost of provisioning and the discharge bounty and are shown to the nearest HS 100 million. Dramatic increases are shown as occurring under Septimius Severus and Caracalla. These increases flow from the increase in base pay rates and also, but to a lesser extent, from increases in the size of the army. The amounts are in HS.⁴⁶⁸

⁴⁶³ Duncan-Jones (1990) 116 for the estimates of annona and donatives. The allowance for wheat is 30 modii per annum and for oil is 1/11 sextarius per day. Duncan-Jones (1990) 106-110.

⁴⁶⁴ This is the equivalent of 13 years' salary for discharge after 20, later 25, years' service.

⁴⁶⁵ See Rathbone (2007a) 163 for a discussion of the bounty.

⁴⁶⁶ The initial endowment was about two years' liabilities.

⁴⁶⁷ See Hopkins (1980) 124.

⁴⁶⁸ The total cost is calculated as average cost per soldier x number of soldiers + discharge bounty. The discharge bounty is calculated as HS 12,000 multiplied by 4% of the number of legionary soldiers. The 4% assumption is broadly consistent with Hopkins (1980). I assume that the removal of deductions that happened in the late second century was equivalent to an increase in base costs of 30%. I make no allowance for the beneficial effect on the state's finances of the increase in deductions made from soldiers pay under Domitian.

Table 8.5: Calculation of the total cost of the army

Assumption	Augustus	Domitian	Septimius Severus	Caracalla	Diocletian
	27 B.C. –AD 14	AD 81 -96	AD 193- 211	AD 198-217	AD 284-305
Base Pay	900	1,200	2,400	3,600	7,200
Higher grade Supplement	270	360	720	1,080	2,160
Annona					2,400
Donatives					40,000
In kind payments			720	1,080	1,440
Average cost per soldier	1,170	1,560	3,840	5,760	53,200
Discharge Bounty	60 million	70 million	70 million	80 million	no longer paid
Total cost of the army	350 million	500 million	1,200 million	2,500 million	22,000 million

Estimates have been made by scholars for total army costs under Augustus, Domitian and Septimius Severus.⁴⁶⁹ My estimates for costs under Augustus and Domitian are generally in line with these estimates but substantially higher for costs under Septimius Severus. The annual recurring cost of the army after the increase given by Caracalla is in excess of the *patrimonium* of Augustus. The following table summarises those estimates. The amounts are

⁴⁶⁹ Rathbone (2009) 311 gives an estimate of average cost per soldier of HS 47,200 under Diocletian, which is not significantly different to the figure I use.

in HS millions ⁴⁷⁰ Note that much of this increase is inflationary rather than a real increase as illustrated by graph 7.6.

Table 8.6

Summary of scholarly estimates of total army costs – excluding discharge bounty

Estimate	Augustus	Domitian	Septimius Severus	Caracalla	Diocletian
Frank	275				
MacMullen		400	576		
Hopkins	379				
Develin		464	619		
Rathbone	500+		850		
My estimates	300	400	1,100	2,400	22,000

In broad terms the cost of the army, including discharge bounty, in the first two centuries AD was about HS 0.5 billion. The increases that took place, very likely starting with Commodus, towards the end of the second century and certainly continuing under Septimius Severus and Caracalla increased this cost to about HS 2.5 billion. The following table sets out the current account of the state.⁴⁷¹ In the early part of the period there are substantial excess revenues.⁴⁷²

⁴⁷⁰ Frank (1940). The figure is for legionaries only. MacMullen (1984) 572. Hopkins (1980) 125, but excluding his allowance for discharge bounty. Develin (1971) 691. Rathbone (2007a) 174, taking his post Domitian figure as being comparable to the Septimius figure and loading both figures by 25% to allow for officer pay.

⁴⁷¹ The inflation of the late second century will have increased state revenues but not to the extent of changing the broad conclusions illustrated by this table.

⁴⁷² In addition to the change from a conquest to a trade model there is also an increase in the ability of the centre to exercise patronage from the provincial revenues. But as noted in section 8.5 there are limits to that patronage, even with the large sums available. The pay increases of the Severans destroyed that power.

Table 8.7 Schematic of State Current Account

	Mid second Century	Septimius Severus AD 197	Caracalla AD 212
Income	1,500	1,500	1,500
Cost of army	500	1,200	2,500
Excess	1,000	300	(1,000)

8.7 Conclusion

The immediate consequence of the decision to substantially increase the soldiers' pay was that the existing system of raising taxes was no longer sufficient to pay the ongoing operations of the state.⁴⁷³ Taxes needed to be broadly doubled or sequestrations needed to become a continuing source of revenue as significant as the provincial tribute or the currency needed to be debased or assets needed to be acquired through conquest or promised payments such as soldiers pay had to be reneged on. All of these options, with the exception of conquest, are destabilising and the potential gains from conquest are, in any case, limited.⁴⁷⁴ Other options included asset sales which could also not be used as a source of recurring revenue, and reductions in expenditure.⁴⁷⁵ The increase under Septimius Severus was at a level which could be broadly managed through these options. The deficit under Caracalla was such that the basic functions of the state, or at least those which required substantial state

⁴⁷³ Duncan-Jones (1994) 15 describes extraordinary sequestration and higher taxes as the means by which Septimius Severus paid for the increased pay of the army.

⁴⁷⁴ If the recurring income is about half of the recurring expenditure then broadly, if the debasement route is the only option exercised, the silver content must be reduced each year by 50% of what it was in the previous year since the underlying deficit remains the same. Debasement in the reign of Commodus was 28% in total, in the reign of Septimius Severus was 13% in total and in the reign of Caracalla was 12%

⁴⁷⁵ There is evidence of some reductions in state expenditure around this time Wilson (2011). Municipal expenditure, however, continued in North Africa, for example.

funding, could no longer be paid for from the normal revenues of the state.⁴⁷⁶ These functions included the army, the provision of the grain dole and provincial government. Other functions, which required little funding, such as the judicial functions could continue unaffected; indeed the third century is the golden age of the jurists. The consequences were severe. The army could no longer be afforded, or at least it could no longer be paid under the remuneration structure of fixed cash amounts paid three times yearly which had been established by Caesar. In the middle of the second century AD the revenue collected by the state had substantially exceeded the cost of the army and there had been a substantial surplus to fund the provincial and central government. The surplus was no longer available from the tax revenues of the empire.

The fifty years that followed the reign of Caracalla was a period of extreme political instability with continuing rapid replacement of Emperors because, in my view, each attempted to solve a problem which could not be solved without significant structural change and was removed after the inevitable failure. Had the increases been of the order of 30%, say, then the disruption would have been minor. The decisions to award increases to the army which were beyond the ability of the state to pay were either reckless or incompetent. They were taken either in recognition of the effects that they would inevitably have or were taken ignorant of those effects. Septimius Severus became Emperor after a career which was exclusively military and with no prior exposure to the processes by which the state was administered and financed.

The increases in pay that took place over the fifteen year period, AD 197 to AD 212, were incompetent and changes in how the leadership of the state emerged may have been a contributory factor.⁴⁷⁷ The career system under the Republic had had a significant meritocratic element and individuals who reached the highest levels in the governance of the state had had to have followed career paths which were such that they had acquired proven practical experience in the areas of military, financial, organisational and judicial matters. Progression to the later stages in the career structure depended on election under the Republic when the candidate's competence or otherwise in the previous roles was available to his rivals as debating material. The process, therefore, tended to lead to the selection of individuals who were literate and capable in the various aspects of government. It is possible

⁴⁷⁶ Crawford (1977) 568 notes that by the time of Caracalla the army was no longer affordable.

⁴⁷⁷ The increase occurred at the same time as the flow of silver from the Spanish mines was reduced.

that this process had been undermined and that the decisions on army pay were made with a failure to appreciate the consequences of such an increase in recurring expenditure.

Chapter 9: Conclusions

9.1 Introduction

In this section I draw together the analyses of the preceding chapters. I first consider the process by which the money system was created and the possibilities that then emerged. I then consider the structure of the economy of the Roman Empire in terms of the separation of town and country, the agglomeration of individuals in the towns and the connections between those towns. I then return to the development of the economy over time with a discussion of the destruction of the money system. Finally, I describe the results of this experiment.

9.2 Creation of the money system

During the late Republic there had been a dramatic increase in the supply of money. This increase had occurred as the state acquired precious metals and produced coins which it used to fund expenditure. The army was the principal agency through which this money was distributed into the territories which were controlled by the state. This level of monetisation continued through the Roman Empire and facilitated the changes that were to happen.

Money is not just a means of exchange, store of value and a unit of account. It also enables the development of economic relationships. The use of money opened up the possibility of a change in the relationship between producer and consumer. With money it was no longer necessary that the two had a relationship based on one producing what the other needed. Goods and services could be produced which were widely separated in time and space from those who ultimately consumed them; the ultimate producers and consumers of these goods and services might be unknown to each other. Transactions could take place in which one large item might be sold to a number of consumers, none of whom individually had need of or the wherewithal to acquire that single item. A farmer could rear cattle for slaughter and sell the meat to men who could not afford and did not need a single animal. This development in economic relationships, starting in the late Republic, followed from the creation of money by the state and its distribution by a state agency together with the free construction of new economic relationships between individuals.

With the wide dispersal of the state coins and the increased depth of monetisation that followed from the continuing distribution of coin into the territories controlled by the state further changes happened. In effect, the only asset class that had been available for investment had been land, and while this asset class continued to support and enhance status

and while investment in this asset class by the elite continued to be glorified, the state had gradually created and expanded a new asset class. This asset class was money. Access to money, through loans, for example, enabled trading enterprises to acquire working capital, and fixed assets, such as ships, and fortunes could be made. Those serving in the army had the opportunity to accumulate funds from the regular salary paid to them and their discharge bounty could be used to fund ventures. The necessary skills to start these ventures were there to be acquired through service in the army. Basic forms of literacy and the rudiments of accounting, visible in the soldiers' pay records, were there to be absorbed as was the knowledge of how different parts of the Roman Empire operated, where things were readily available and where they might be needed. The monetisation of the economy and the development of money as an asset class enabled these skills and this knowledge to be applied and the only limitations were the desire and the ability of the individual to do so. The story of Trimalchio is, in my view, a portrayal of the nervousness felt by the group for which Petronius writes and which derived its wealth from land and the service of the state, as it sees the wealth that can be generated in the new system and the new power to acquire goods and services that they also want.

The monetary system that was being constructed was metallic and for its smooth operation required that there be fixed relationships between the different denominations that were issued by the state. It required that there be trust in the continuing acceptability of these coins, which was partly based on their actual metal content, on the state's willingness to accept them back as payment of tax and also, in time, on the long-customary acceptance of the coinage. As discussed in Chapter 7, the currency had the weakness that all metallic coins had until the start of the nineteenth century, which lay in the arbitrage opportunities which encouraged currency destruction. This risk threatened the availability of money as an asset class and it threatened trade. State actions, and the wide expanse of the territory controlled by the state, were, at least for a period, protection against these risks. When Vespasian centralised the mint and when rival currencies were removed from the system the coinage was protected by the demand that was further sustained for the currency of the Roman Empire through the elimination of currency competition.

There were further weaknesses in this money system. Money was created out of precious metals and not, as in modern systems, out of thin air and the stock of money in circulation was continuously depleted through the continuing chance loss of coin and the export of coin from the Roman Empire into overseas territories, such as India. So long as the reserves of the

state were adequate money could be continued to be created. These reserves derived from the treasures acquired in conquest, the flow of which decreased as the conquest opportunities subsided, and metals taken from the mines which the state controlled, for example. A significant and positive feature of this economic structure, however, is that the controllers of assets such as these used them to create money which they distributed, through the army pay structure, building projects and civil extravagances which were all part of the mechanism of money creation. Alternative uses of these metals had been possible and were applied to an extent, such as the accumulation of temple treasures and the export of coin from the state in exchange for imports, but to a very great extent the metals were applied in a process of increased monetisation which furthered economic activity.

In Chapter 7 I showed how the currency that was constructed was stable through the first and second centuries AD. Prices of goods expressed in terms of the coins of this currency were stable, apart from the normal variations to be expected because of changes in the supply or demand for particular goods, as for example when weather conditions reduced the supply of grain. The relationship between the different denominations, particularly between those struck in gold and in silver, was broadly stable and the currency was accepted over a wide area. This period of currency stability, in all these aspects, is probably unprecedented.

9.3 Separation, agglomeration and connection

Grain has the property that it does not deteriorate over a reasonable period of time when stored properly, and therefore enables the separation of the points of production and the points of consumption. This in turn enables people to live in centres of high population density and changes the nature of the human interaction. Close proximity increases the contact between individuals and facilitates the dissemination of knowledge.

Agricultural productivity, at least from the end of the Roman Republic, was such that levels of urbanisation could be achieved which were not reached again until around the middle of the eighteenth century in England, as discussed in Chapter 3. Unremarkable towns of imperial Rome exceeded in size anything that was achieved in medieval England, for example. As shown in Chapter 2, this level of grain productivity was not a consequence of any technological difference but because of difference in work practices. At its most basic the time and effort put into the cultivation of the crop, in particular into weeding, made a significant difference. I attribute the difference to the fact that land in the time of imperial Rome was a marketable asset in a way in which it was not in feudal systems. This

marketability was enhanced by property valuation approaches which were sophisticated and by the creation of derivatives constructed from property assets, which were in turn capable of valuation and could be transferred.

High levels of urbanisation require large agricultural surplus which in turn requires high agricultural productivity but also efficient labour structures. The strategies employed on landholdings are determined by the controller of production, within the limits of what is permitted within the economy. There were few restrictions on the strategies that might be employed unlike in, for example, medieval England and efficient labour structures such as slave estates or consolidated land holdings, for example, were constructed. The increased level of monetisation facilitated the construction of these estates. The output could be sold off the estate for money and paid for by money earned in urban occupations. The increasingly plentiful supply of money facilitated these developments. The process of constructing an efficient rural landscape through consolidation is also a process in which labour migrates to the urban centres. Small holdings which do not provide an outlet for all of the labour of those who live on them are consolidated with other small holdings. Some of those who had previously lived on the land leave and those who remain are more fully employed.

The depopulation of the rural landscape contributed to the growth of the towns and cities, as it was to do later in nineteenth century England. This depopulation contributed to agricultural surplus since the same amount of food was produced by a smaller rural population. The sizes of the cities that emerged, a number having populations in excess of 500,000, was such that the economic relationships required money to function. Without the high level of monetisation that had emerged life in these urban centres was not viable. Urban centres of this size also require substantial infrastructure for the provision of food, water and the removal of waste.

The cities were freely connected through a trade network across the Mediterranean. This trade network depended on the state and the local municipalities for infrastructure creation and maintenance but was otherwise a private enterprise. The economic structure that had evolved was a network of highly connected urban centres spread across a wide area. These urban centres were embedded in an agricultural landscape which was poorly connected for the transportation of goods. In Chapter 3 I discussed how agricultural land more than about ten kilometres from an urban centre was generally self-sufficient estates or small agricultural units producing little beyond what the inhabitants needed. The network was supported by a

common currency, law and language and there were few hindrances to the free flow of goods around this network. This structure had further consequences for agricultural productivity since the output from fertile land could find markets across the network and was not restricted to local markets, provided that it was within reach of connections into the trade network. Average agricultural productivity became more weighted to that achieved on good land.

Where urban populations are more productive than rural populations, because of the types of occupation which are available in urban settings and because of the opportunities to specialise, then increases in urbanisation lead to increases in total output. The very high levels of urbanisation evidenced in the Roman Empire when compared with that seen in medieval Europe suggest that the total output per capita experienced in the time of the Roman Empire was much higher than in medieval Europe. This high level of output was facilitated by the infrastructure which made these large cities possible and by the spreading of the technology required to generate such infrastructure through the empire.

I believe that total output increased further during the second century AD. Through this period there was significant increase in the infrastructure that was available to support trade. As I discussed in Chapter 4, the processing capacity around Rome increased roughly fivefold. Since ships leaving one port need to be processed at another this increase at Rome suggests more general increase in processing capacity and there is evidence of this in construction in North Africa, for example. The percentage increase in processing capacity at Rome is broadly comparable to that which occurred in the world from the middle of the twentieth century. That increase happened not because of changes to the land sea interface but of the great reduction in processing time that occurred when cargo started to be carried in containers rather than in loose sacks or barrels. The recent rise of China as an economic power is attributable to the increase in port-processing capacity.

The level of trade increase that happened in the Roman Empire through the second century AD is such that we expect further improvements in productivity because of specialisation and the exploitation of comparative advantage. These volumes of trade can also only have been economic if wide markets were being accessed and suggest that the non-elite consumers played an increasingly important part in the process. This economic development further increased per capita total output beyond that which was later produced in medieval times. It is

likely, in my view, that per capita total output reached levels next seen in eighteenth century England, or later.

The connectivity which enabled this improvement in material culture flows directly from the building of infrastructure under such emperors as Trajan. This is discussed in Chapter 4 and in Chapter 6. The correspondence between Trajan and Pliny shows how the administrative structures facilitated the dispersal of the knowledge needed to generate these structures and how the elite group of which they were members saw its status as being enhanced by this activity which furthered economic development. Investment, however, is required not just to create these structures but also to maintain them. If the buildings used to temporarily store goods carried into a port are not maintained, or if rivers used for interconnection are allowed to silt up, or if officials needed for processing are no longer present in the full quota, or if damage to harbour structures are not repaired, then processing capacity falls. The level of material culture that had been generated by the second century AD depended for its continuance of the continuing application of the capabilities which had created it.

9.4 Destruction of the money system

In Chapter 8 I showed that when Septimius Severus increased the pay of the army, at the end of the second century AD, and then Caracalla did so again, they began a process which fundamentally altered the dynamics of the money system. The difficulty that they created flowed from the fact that their decision was irreversible, because it was an increase in the regular pay of the army. Indeed, later attempts by Macrinus to reduce the cost of the army failed. Septimius Severus and Caracalla imposed on themselves and their successors the obligation to continue to pay the rewards which they had granted. This doubling of the largest recurring item of state expenditure led to an increase which was beyond what could be financed from the regular tax revenues of the state. The cost of the army had been HS 350 million under Augustus. The tax revenues of the state had been HS 1,500 million. The cost of the army had increased to HS 500 million under Domitian and this was increased by Septimius Severus to HS 1,200 million. The large increases which began with Septimius Severus continued through the third century AD and led to a cost of the army of HS 22,000 million by the time of Diocletian.

These changes led to a fundamental alteration in the army pay structure, which had been constructed under Caesar. That pay structure had consisted of three payments each year of the same amount and from his income the soldier had to meet the cost of such items as food and

equipment. A 'pension' arrangement had been put in place by Augustus. This pay structure had been unwound through the third century and by the time of Diocletian the bulk of the financial benefits that the soldier received was in the form of donatives paid on the anniversaries of the accession of the Augusti and the Caesars. An orderly system of regular state pay had been replaced by one of personalised donations.

The cost of the army had been a significant part of state expenditure but had, at least in the early Empire, been much less than half of the regular revenue of the state. The fact that any increase in expenditure, because of warfare or imperial extravagance, was often associated with extraordinary measures such as the debasement of the coinage suggests that the excess of state revenues over the cost of the army was normally spent and that the state did not run regular surpluses. The cost of the operation of the state, as measured by the difference between the revenues and the cost of the army, was not small. The sudden and dramatic increases put in place by Septimius Severus and his successor, Caracalla, were highly disruptive. They took place at the time when the resources used by the state to create money were under great pressure as the output of the Spanish mines declined. The payment of the cost of the army and other recurring items of state expenditure had to be financed through fund raising which included disorderly mechanisms such as sequestration, debasement of the coinage or extraordinary increases in taxation.

The money system that was in place at the beginning of the third century AD had been in operation for several hundred years. Periodically and routinely as the state expenditure exceeded its revenues new money had been created. There had been reductions in money supply from the simple loss of coin, the melting down to make jewellery and the transfer of coin out of the state into other territories. The effect of this process of money creation over a long period of time was that a very large volume of silver was dispersed in coin across the broad area of the Roman Empire. This money was the basis for commercial transactions, and supported urbanisation and trade. This system was now undermined.

The silver content of the coinage had always varied from time to time but the debasement became severe starting in the beginning of the third century AD. The operation of Gresham's law means that coins with higher silver content will tend to be withdrawn and the more debased coins will remain in circulation. A system under which large volumes of silver circulated in the form of coin was now under pressure and silver was withdrawn at a time when the resources by which the state might replace that silver became limited. The

mechanisms which in the past had been used to collect silver in an orderly way from the population though taxation were increasingly collecting base metal. The only mechanism by which the state could now acquire precious metal was by sequestration.

In Chapter 7 I showed that by the middle of the third century the price stability which had prevailed through the first two centuries of the Roman Empire began to disappear. By the time of Phillip the Arab the fixed relationship between the gold and the silver coins, which had been an essential feature of the currency, had also disappeared. Infrastructure development had largely stopped. Towards the end of the third century hyperinflation had set in. The army was now largely paid in kind.

9.5 The experiment

Weber believed that *homo oeconomicus* first emerged in the Middle Ages. I do not believe that this construct of a man who is motivated purely by the maximisation of profit applies widely even in modern times, when status matters and can influence behaviours in commercial transactions as much as the mere opportunities for monetary gain. However, I believe that during the time of imperial Rome there were opportunities for men who did not belong to the land owning elite to create large fortunes, that by the start of that period the members of the land-owning elite no longer had a monopoly on wealth creation and that as trade increased, and with it total output, the contribution of land declined proportionately. The ability and willingness to negotiate, which Weber saw as contrary to the values of the elite group in Antiquity, is evident in the behaviours of the elite in the time of imperial Rome and in any case underlies trade which was extensive in the first and second centuries of the Roman Empire. There is not much difference evident in the commercial behaviours of the men of business of that time and of modern times and much that is similar. To the extent that *homo oeconomicus* exists he was to be widely found in the Roman Empire as much as in the Middle Ages and later.

I believe on the evidence of agricultural productivity, the level of urbanisation, the extent of trade and the life expectancy of the population that economic output as measured by the quantity and quality of the goods and services produced in the first and second centuries AD of the Roman Empire was broadly equal to that achieved in the middle of the eighteenth century in England, when adjusted for differences in population size, and was probably only reached much later by other countries. The level of economic output of the Roman Empire in that period was substantially above that produced in medieval Europe, again adjusted for

differences in population size. The level of economic output of the Roman Empire is very likely to have declined from at least the middle of the third century AD as the currency became unstable and as infrastructure spending reduced.

The institutional structures were such that output was maximised relative to the technology which was available. Significant developments which did not occur and which could have led to further improvements in output were the steam engine - which would have increased the land-based hinterland of towns, brought further areas of cultivatable land within the reach of urban centres and increased the levels of maximum possible urbanisation - and changes to the genetic material of wheat, in particular the shortening of the stalk. Other changes in the knowledge base, such as those which emerged from the seventeenth and into the nineteenth centuries and which led to the development of new forms of material goods were very likely too far beyond reach.

The long period of peace over a wide geographic area and, in particular, the freeing of the Mediterranean from piracy, supplemented by a widely accepted and stable currency provided the ideal conditions in which economic development could happen. The institutions which supported that peace contributed significantly to the economic achievement as did the institutions which maintained land, and property more generally, as a marketable asset and which allowed the free movement of people between urban centres across the empire. The widespread construction of infrastructure necessary for urbanisation and for trade, which was supported by local and central institutions, facilitated the development.

Appendix 1

Relationship between agricultural production and urbanisation.

Definitions:

- a) Let **R** be the number of rural workers and their dependents. Together they each produce a grain surplus of **S**. Surplus **S** is measured in units of a man's food requirement.
- b) Let **U** be the number of productive urban workers and their dependents. Together they each produce goods or services of **G**. Goods **G** are measured in units of a man's food requirement.
- c) Let **C** be the number of urban pure consumers.
- d) Let **u** be the urbanisation percentage, the percentage of the population which lives in the urban centres.

By urban dwellers I mean inhabitants of areas exclusively populated by people who produce scarcely any of the food they consume. The unit of measurement is the annual food consumption of a man.

Relationship:

urban food consumption = agricultural surplus

$$C+U = S.R$$

urbanisation percentage = total urban dwellers/total population

$$u = (C+U)/(R+C+U)$$

$$= S.R/(R+S.R)$$

$$= S/(1+S)$$

This level of urbanisation is the maximum that can be sustained by a system in which surplus **S** is produced. In the configuration described above all the grain surplus is consumed in the urban centre. Other configurations are possible in which surplus **S** is exchanged with rural units which do not produce grain. The urban population in these other configurations is lower because only a proportion of the surplus grain is consumed in the urban centre.

Examples:

- a) Assume that on average 10 rural dwellers produce enough surplus to support 1 urban dweller. Rural dwellers are all those who live on the land, whether economically active or not. Then the proportion of the population which lives in urban centres is $1/(1+10) = 9\%$
- b) Assume that on average 2 rural dwellers produce enough surplus to support 1 urban dweller.

The proportion of the population which lives in urban centres is $1/(1+2) = 33\%$

Observation on the parameter S:

S is a measure of urbanisation since if

$$u = S/(1+S) \text{ then}$$

$$S = u/(1-u)$$

$$= \text{number of urban dwellers/number of rural dwellers}$$

Appendix 2

Imperial Roman land based transportation parameters

This subsection considers the parameters that describe the available land based transport systems. These are the range that method of transport could cover in a journey, the speed at which it could move, load it could bear and the hire cost. These variables are more deterministic than the variables which underlie seed productivity, which are heavily stochastic, and the quantification is less problematic.

On land grain was transported by men, by pack animals and by wheeled transport drawn by animals. Pack animals did not need any infrastructure and could go over rough tracks. Animals pulling wheeled vehicles needed some sort of road structure but the infrastructure required for bullock wagons was relatively basic. Cotterall and Kamminga provide an analysis of the key parameters for this transport system, which are summarised in the following table.⁴⁷⁸

Table A.2.1 Load bearing capacity and range of different transport types

1: Form of transport	2: Maximum load	3: Daily range
Man	50 kg	20 km
Donkey/mule	150 to 180 kg	20 to 24 km
Wheeled – per animal	680 kg	29 km

Pack animals had somewhat shorter daily ranges than wheeled transport and carried significantly lighter loads than wheeled transport. These means of transport generally moved at rates of about 4km per hour. The normal journey involved leaving and returning on the same day and the range of the land based distribution system was therefore about 10 to 15 kilometres. This was effectively the maximum radius of the hinterland of towns and villages.

The main arterial highways were state constructions; the building of major roads was a political activity. There had been substantial construction during the Republic and this

⁴⁷⁸ Cotterall and Kamminga (1990) representative figures except range man.

continued into the Principate. Claudius built 320 km of roads and there was significant building during the time of Trajan. These roads were of very high quality and linked the main centres of population. In addition to these highways there were subsidiary roads which were built by local municipalities and which fed into the consular network. Local municipalities also built roads linking towns. Private roads were constructed by estate owners across their properties and to link into the local and consular networks.⁴⁷⁹ This network facilitated the smooth movement of the army, the flow of government information and the movement of people and goods. However, the difficulties of moving bulky or heavy goods through this system are widely appreciated.⁴⁸⁰

The inter-centre distances in this network range from 3 to 64 km. About 60% of the towns and villages in Italy were between 25 km and 40 km distant from another settlement. The average staging between journeys was about 30 km.⁴⁸¹ Land transport was supplemented by movement along rivers which was considerably cheaper and with a greater range. Movement upstream was, however, difficult in some cases.⁴⁸²

⁴⁷⁹ Laurence (1999) 5-53.

⁴⁸⁰ Rickman (1980) 13 and Russell (2009) 110 note the difficulties of road transport.

⁴⁸¹ Laurence (1999) 86-90.

⁴⁸² Russell (2009) 112-4.

Appendix 3

Relationship between urban percentage, relative urban productivity and total economic output

Definitions:

- a) Let O_r be the rural output per capita measured in monetary units and calculated as total rural output divided by total rural population.
- b) Let O_u be the urban output per capita measured in monetary units and calculated as total urban output divided by total urban population.

Relationship:

$$\begin{aligned}
 \text{total output} &= \text{rural output} + \text{urban output} \\
 &= R.O_r + (C+U).O_u \\
 &= R.O_r.(1+(C+U)/R. O_u /O_r) \\
 &= \text{rural output}.(1+(1/(1-u)-1)) \text{ .relative urban productivity)} \\
 &= \text{rural output}. (1+S.\text{relative urban productivity})
 \end{aligned}$$

This can be approximated as

$$\text{total output} = \text{food produced} \times (1 + \text{urbanisation proportion} \times \text{relative urban productivity})$$

under the simplifying assumptions that only food is produced outside the urban centres and that no food is produced within the urban centres. This is to ignore mining, for example, and the production of clothes and agricultural tools for own consumption. Urbanisation proportion is the urban population as a proportion of the rural population.

Examples:

- a) The urbanisation percentage is 9% and urban worker is twice as productive as the rural labourer. Total output is rural output *1.1
- b) The urbanisation percentage is 33% and the urban worker is four times as productive as the rural worker. Total output is rural output *3.0

Appendix 4

The silver content of the denarius

Table A4.1 below shows the silver content of the denarius, or its equivalent, for the period AD 64 to AD 274. The grams of silver shown are those in the denarius for the period AD 64 through AD 241. The grams of silver show for the period AD 243 through AD 270 are half what was in the antonianus, which is taken as being equivalent to two denarii. The grams of silver shown for AD 274 are one fifth of what was contained within the aurelianus which is taken as being equivalent to five denarii.

The data is taken from <http://www.tulane.edu/~august/handouts/601cprin.htm>. Retrieved 20 October 2015. These figures were compiled by K. Harl from data in Walker (1976), Walker (1977) and Walker (1978). Walker's figures were based on analyses of coins using x-ray fluorescence spectroscopy of the metal content of the surface of the coin. This method has the advantage that it is non-destructive.⁴⁸³ For the later coinages, the coins are so plentiful that destructive methods are possible.⁴⁸⁴

Table A4.1 Silver content of denarius or its equivalent: Walker estimates

1: Period AD	2: Grams of silver
64-68	2.97
70-81	2.87
82-85	3.26
85-107	3.04
107-148	2.88
148-161	2.68
161-168	2.57
168-170	2.67
170-180	2.57
180-185	2.34
186-192	2.22

⁴⁸³ Walker (1976) 1-3.

⁴⁸⁴ Walker (1978) 1.

193	2.75
193	2.40
193-194	2.46
194-196	1.98
196-211	1.81
212-217	1.66
217-218	1.82
219-222	1.41
222-228	1.30
229-230	1.46
230-235	1.50
235-238	1.43
238	1.71
238	1.55
241	1.46
243	0.81
244	0.87
248	0.97
250	0.82
251	0.63
253	0.34
255-260	0.28
260	0.27
261-263	0.23
263-265	0.19
265-266	0.16
267-268	0.13

268	0.05
269	0.03
270	0.05
270	0.04
274	0.04

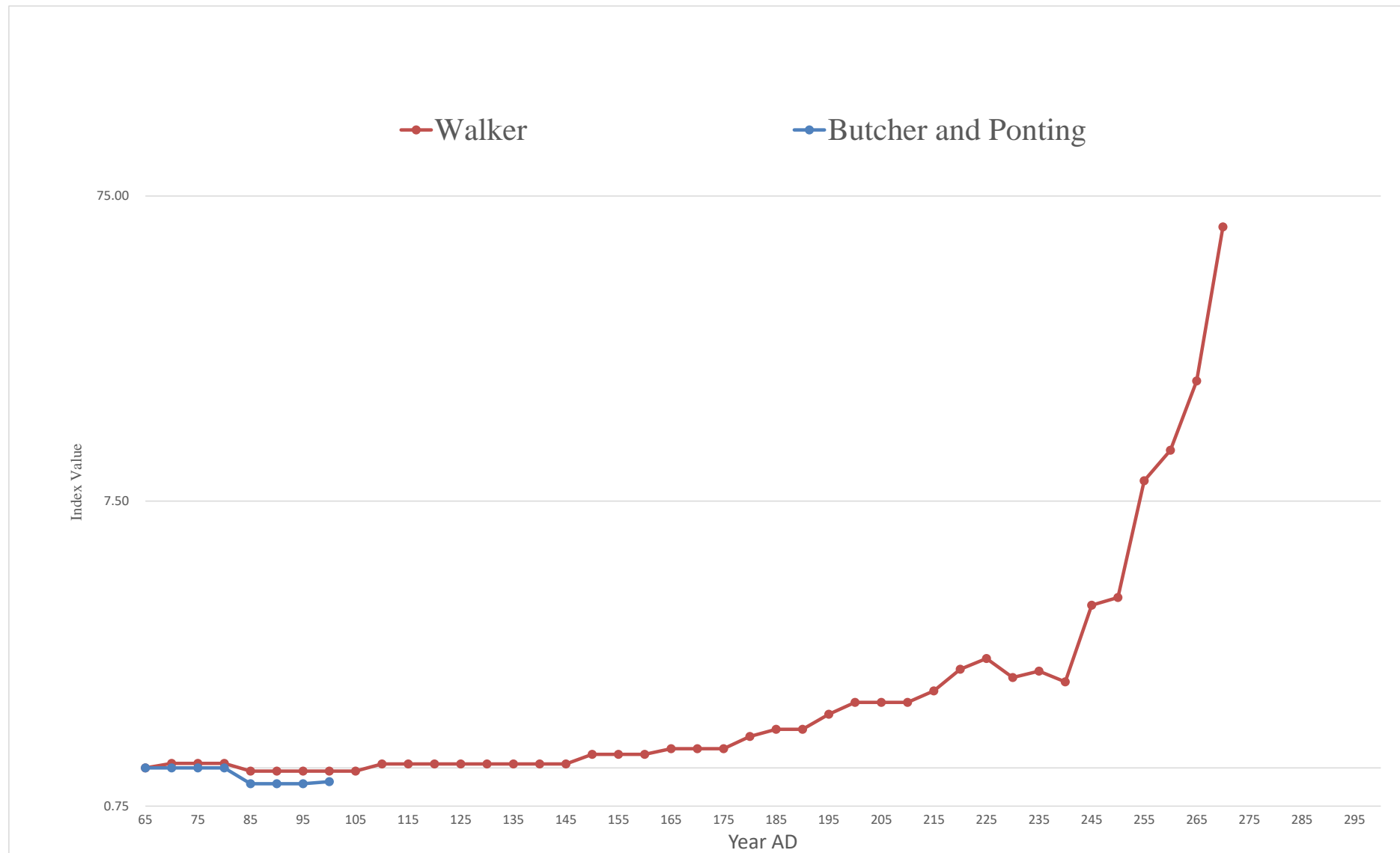
A potential disadvantage of the method used by Walker is that it only examines the silver content of the surface of the coin and there is the risk that the mint deliberately sought to deceive by producing coins with a higher silver content on the surface and a lower silver content in the core of the coin. Butcher and Ponting used inductively coupled plasma atomic emission spectroscopy to carry out a second analysis of denarii. The following table shows their estimates of the silver content of the denarius.⁴⁸⁵

Table A4.2 Silver content of the denarius: Butcher and Ponting estimates

1: Emperor	2: Period AD	3: Grams of silver
Nero	54-64	3.65
Nero	64-68	2.76
Nero	68	3.11
Glaba	68-69	3.11
Otho	69	3.11
Otho	69	2.72
Vitellius	69	2.76
Vespasian	69-79	2.72
Titus	79-81	2.76
Domitian	81-82	2.72
Domitian	82-85	3.55
Domitian	85-96	3.11
Nerva	96	3.11
Nerva	97-98	2.97
Trajan	98-99	3.06

⁴⁸⁵ Butcher and Ponting (2014) 701.

The following graph compares two indices of silver content constructed from these two sets of estimates.

Figure A4.1 Comparison of Walker data with Butcher and Ponting data

- Allen, R (2009), 'How prosperous were the Romans? Evidence from Diocletian's Price Edicts (A.D. 301)', in A. Bowman and A. Wilson (eds.), *Quantifying the Roman Economy – Methods and Problems*, Oxford: 327-45.
- Alston, R. (1994), 'Roman military pay from Caesar to Diocletian', *Journal of Roman Studies* 84: 113-23.
- Alston, R. (2007), 'Warfare and the state - the military and politics', in P. Sabin, H. Van Wees and M. Whitby (eds.), *The Cambridge History of Greek and Roman Warfare*, Cambridge: 158-177.
- Andreau, J. (2010), *L'économie du Monde Romain*, Paris.
- Arnaud, P. (2007), 'Diocletian's Prices Edict: the prices of seaborne transport and the average duration of maritime travel', *Journal of Roman Archaeology* 20: 321-36.
- Austin, R., M. Ford and C. Morgan (1989), 'Genetic improvements in the yield of winter wheat: a further evaluation', *The Journal of Agricultural Science* 112 3: 295-301.
- Bagnall, R. and B. Frier (1994), *The Demography of Roman Egypt*, Cambridge.
- Bailey, D. (1996), 'Honorific columns, cranes and the Tuna epitaph', in Bailey, D. (ed.), *Archaeological Research in Roman Egypt*, Ann Arbor: 155-68.
- Bang, P. (2007), 'Trade and empire: in search of organizing concepts for the Roman economy', *Past and Present* 195: 3-54.
- Barclay, W., A. Coale, M. Stato and T. Trussell (1976), 'A reassessment of the demography of traditional rural China', *Population Index* 42 4: 606-635.
- Barsby, A. and C. Barsby (1996), *Manorial Law*, Epsom.
- Bastianelli, S. (2008), *Centumcellae (Civitavecchia), Castrum Novum (Torre Chiaruccia): Regio VII, Etruria*, Rome.
- Blayo, Y. (1975), 'La mortalité en France de 1740 à 1829', *Population* 30: 123-42.
- Bloom, D., D. Canning and G. Fink (2008), 'Urbanisation and the wealth of nations', *Science* 319: 772-5.
- Blum A. and I. Troitskaya (1997), 'Mortality in Russia during the 18th and 19th centuries: local assessments based on the Revizii', *Population: An English Selection* 9: 123-46.
- Broadberry, S., B. Campbell, M. Overton, B. van Leeuwen and A. Apostolides (2009), 'Historical National Accounts for Britain, 1300-1850: some preliminary estimates', *Reconstructing the National Income of Britain and Holland, c. 1270.1500 to 1850*.
- Brunt, P. (1950), 'Pay and superannuation in the Roman army', *Papers of the British School at Rome* 18: 50-75.
- Brunt, P. (1981), 'The revenues of Rome', *Journal of Roman Studies* 71: 161-72.

- Burn, A. (1953), 'Hic breve vivitur: A study of the expectation of life in the Roman Empire', *Past & Present* 4: 2-31.
- Burnett, A., M. Amandry and P. Ripollès (1992), *Roman Provincial Coinage. Volume I: From the Death of Caesar to the Death of Vitellius (44 BC-AD 69)*, London.
- Butcher, K. and M. Ponting (2014), *The Metallurgy of Roman Coinage: From the Reform of Nero to the Reform of Trajan*, Cambridge.
- Cairncross, A. (1949), 'Internal migration in Victorian England', *Manchester School* 17: 67-87.
- Cairncross, A. (1953), *Home and Foreign Investment*, Cambridge.
- Carroll, M. (2014), 'Mother and infant in Roman funerary commemoration', in M. Carroll and E.-J. Graham (eds.), *Infant Health and Death in Roman Italy and Beyond*, *Journal of Roman Archaeology Supplementary Series*: 159-78.
- Casson, L. (1965), 'Harbour and river boats of ancient Rome', *Journal of Roman Studies* 55: 31-9.
- Casson, L. (1995), *Ships and Seamanship in the Ancient World*, 2nd edn., Baltimore.
- Champlin, E. (1991), *Final Judgements: Duty and Emotion in Roman Wills 200 B.C. to A.D. 250*, California.
- Coale, A. and P. Demeny (1983), *Regional Model Life Tables and Stable Populations*, New York.
- Collins, H. (2001), 'Tacit knowledge, trust and the Q of sapphire'. *Social Studies of Science* 31: 71-84.
- Cotterall, B. and J. Kamminga (1990), *Mechanics of Pre-Industrial Technology*, Cambridge.
- Crawford, M. (1970), 'Money and exchange in the Roman world', *Journal of Roman Studies* 60: 40-8.
- Crawford, M. (1975), 'Finance, coinage and money from the Severans to Constantine', in J. Vogt, H. Temporini and W. Haase (eds.), *Aufstieg und Niedergang der römischen Welt: Geschichte und Kultur Roms im Spiegel der neueren Forschung* Berlin: 560-93.
- Crawford, M. (1974), *Roman Republican Coinage*, Cambridge.
- Cuomo, S. (2011), 'A Roman Engineer's Tales', *Journal of Roman Studies* 101: 143-65.
- Davies, R. (1971), 'The Roman military diet', *Britannia* 2: 122-42.
- de Callatäy, F. (2005), 'The Graeco-Roman economy in the super long-run: Lead, copper and shipwrecks', *Journal of Roman Archaeology* 18: 361-72.
- Dessau, H. (1908), 'Afrikanische Munizipal – und afrikanische Militärschrift', *Klio*: 8: 57-63.

- de Vries, T. and W. Zwalf, (2003), 'Roman actuarial science and Ulpian's life expectancy table', in L. de Ligt, E. Hembrijck and H. Singor (eds.), *Roman Rule and Civic Life*, Amsterdam: 277-98.
- Develin, R. (1971), The army pay rises under Severus and Caracalla and the *annona militaris*', *Latomus* 30: 687-95.
- Dixon, J., L. Nalley, P. Kosina, R. La Rovere, J. Hellin and P. Aquino (2006), 'Adoption and economic impact of improved wheat varieties in the developing world', *The Journal of Agricultural Science* 144 6: 489-502.
- Duncan-Jones, R. (1978), Pay and numbers in Diocletian's army', *Chiron* 8: 541-60.
- Duncan-Jones, R. (1982), *The Economy of the Roman Empire*, 2nd edn., Cambridge.
- Duncan-Jones, R. (1990), *Structure and Scale in the Roman Economy*, Cambridge.
- Duncan-Jones, R. (1994), *Money and Government in the Roman Empire*, Cambridge.
- Edgerton, D (2008), *Shock of the Old: Technology and Global History since 1900*, 2nd edn., London.
- Edmondson, J. (1989), 'Mining in the later Roman empire and beyond: continuity or disruption?', *Journal of Roman Studies* 70: 84-102.
- Edmondson, J. (2005), 'Family relations in Roman Lusitania: Social change in a Roman province?', in M. George (ed.), *The Roman Family in the Empire: Rome, Italy and Beyond*, Oxford: 183-229.
- Erdkamp, P. (2001), 'Beyond the limits of the 'Consumer City'. A model of the urban and rural economy in the Roman world', *Historia*: 332-56.
- Erdkamp, P. (2005), *The Grain Market in the Roman Empire – a Social, Political and Economic Study*, Cambridge.
- Evans, J. (1981), 'Wheat production and its social consequences in the Roman world', *Classical Quarterly New Series* 31: 428-42.
- Fifoot, C. (1932), *English Law and its Background*, London.
- Finley, M. (1965), 'Technical innovation and economic progress in the ancient world', *The Economic History Review* 18: 29-45.
- Finley, M. (1977), 'The ancient city: from Fustel de Coulanges to Max Weber and beyond', *Comparative Studies in Society and History* 19: 305-27.
- Finley, M. (1999), *The Ancient Economy*, 2nd edn., Cambridge.
- Foxhall, I. and H. Forbes (1982) 'Sitometreia: the role of grain as a staple food in antiquity', *Chiron* 12: 41-90.
- Frank, T. (1940), *An Economic Survey of Ancient Rome* 5 Volumes, New York.

- Friedman, M. (1990), 'Bimetallism revisited', *Journal of Economic Perspectives* 4 85-104.
- Frier, B. (1982), 'Roman life expectancy: Ulpian's evidence', *Harvard Studies in Classical Philology* 86: 213-51.
- Frier, B. (1994), 'Natural fertility and family limitation in Roman marriage', *Classical Philology* 89 4: 318-33.
- Frier, B. (2000), 'Demography', in A. Bowman, P. Garnsey and D. Rathbone (eds.), *The Cambridge Ancient History XI: The High Empire, A.D. 70-192*, 2nd edn., Cambridge: 787- 816.
- Frier, B. and D. Kehoe (2007), 'Law and economic institutions', in W. Scheidel, I. Morris and R. Saller (eds.), *The Cambridge Economic History of the Greco-Roman World*, Cambridge: 113-43.
- Fulford, M. (2009), 'Approaches to quantifying Roman trade: Response' in A. Bowman and A. Wilson (eds.), *Quantifying the Roman Economy: Methods and Problems*, Oxford: 250-8.
- Garnsey, P. (1983), 'Grain for Rome', in P. Garnsey, K. Hopkins and C. Whittaker (eds.), *Trade in the Ancient Economy*, London: 118-31.
- Gilliam, J. (1952), 'The minimum subject to the *Vicesima Hereditatium*', *American Journal of Philology*, 73 4: 397-405.
- Goldsmith, R. (1984). 'An estimate of the size and structure of the national product of the early Roman Empire', *Review of Income and Wealth* 30: 263-88.
- Gowland, R., A. Chamberlain and R. Redfern (2014), 'On the brink of re-evaluating infanticide and infant burial in Roman Britain', in M. Carroll and E-J. Graham (eds.), *Infant Health and Death in Roman Italy and Beyond*, *Journal of Roman Archaeology Supplementary Series*: 69-88.
- Graham, E-J. and M. Carroll, 'Introduction', in M. Carroll and E-J. Graham (eds.), *Infant Health and Death in Roman Italy and Beyond*, *Journal of Roman Archaeology Supplementary Series*: 9-22.
- Greene, K. (1986), *The Archaeology of the Roman Economy*, London.
- Greenwood, M. (1940), 'A statistical mare's nest?', *Journal of the Royal Statistical Society* 103: 246.
- Gregg, S. (1988), *Foragers and Farmers: Population Interaction and Agricultural Expansion in Prehistoric Europe*, Chicago.
- Halstead, P. (1987), 'Traditional and ancient rural economy in Mediterranean Europe: Plus ça change?', *Journal of Hellenic Studies* 107: 77-87.

- Harari, Y. (2014), *Sapiens: A Brief History of Humankind*, London.
- Harl, K. (1996), *Coinage in the Roman Economy, 300 BC to AD 700*, Baltimore.
- Harris, W. (2006), 'A revisionist view of Roman money', *Journal of Roman Studies* 96: 1-24.
- Harris, W. (2009), 'Comment on Andrew Wilson: 'Approaches to quantifying Roman trade'' in A. Bowman and A. Wilson (eds.), *Quantifying the Roman Economy: Methods and Problems*, Oxford: 259-68.
- Harris, W. (2011), *Rome's Imperial Economy*, Oxford.
- Heinzelmann, M. (2008), 'Supplier of Rome or Mediterranean marketplace? The changing economic role of Ostia after the construction of Portus in the light of new archaeological evidence', *Bollettino di Archeologia*: 5-10.
- Hicks, J. (1969), *A Theory of Economic History*, Oxford.
- Hinde, A. (2003), *England's Population: A History since the Domesday Survey*, London.
- Hinde, A. (2004), 'The use of nineteenth-century census data to investigate local migration', *Local population studies* 73: 8-28.
- Holden, P. (1980), *Studies in the Auxilia of the Roman Army from Augustus to Trajan*, Oxford.
- Holleran, C. (2012), *Shopping in Ancient Rome*, Oxford.
- Hong, S., J.-P. Candelone, C. Patterson and C. Boutron (1994), 'Greenland ice evidence of hemispheric lead pollution two millennia ago by Greek and Roman civilisations', *Science New Series* 265 5180: 1841-3.
- Hopkins, K. (1966), 'On the probable age structure of the Roman population', *Population Studies: A Journal of Demography* 20 2: 245-64.
- Hopkins, K. (1978), *Conquerors and Slaves*, Cambridge.
- Hopkins, K. (1980), 'Taxes and trade in the Roman Empire (200 B.C. - A.D. 400)', *Journal of Roman Studies* 70: 101-25.
- Hopkins, K. (1983), 'Models, ships and staples', in P. Garnsey and C. Whittaker (eds.), *Trade and Famine in Classical Antiquity*, Cambridge: 84-109.
- Hopkins, K. (1987), 'Graveyards for historians', in F. Hinard (ed.), *La Mort, les Morts et au-delà dans le Monde Romain. Actes de Colloque de Caen 20-22 Novembre 1985*, Caen: 113-26.
- Horden, P. and N. Purcell (2000), *The Corrupting Sea*, Oxford.
- Houston, G. (1988), 'Ports in perspective: Some comparative materials on Roman merchant ships and ports', *American Journal of Archaeology* 92: 553-64.

- Howgego, C. (1990), 'Why did the ancient states strike coins?', *Royal Numismatic Society* 150: 1-25.
- Howgego, C. (1992), 'The Supply and Use of Money in the Roman World 200 B.C. to A.D. 300', *Journal of Roman Studies* 82: 1-31.
- Howgego, C. (1994), 'Coin circulation and the integration of the Roman economy', *Journal of Roman Archaeology* 7: 5-21.
- Humphries, A. and R. Biffen (1907), 'The improvement of English wheat yields', *The Journal of Agricultural Science* 2 1: 1-16.
- Johnson, A. (1959), *Roman Egypt: to the Reign of Diocletian*, Baltimore.
- Jones, A.H.M. (1953), 'Inflation under the Roman Empire', *The Economic History Review* 5 3: 293-318.
- Jones, A.H.M. (1964), *The Later Roman Empire*, Oxford.
- Jones, A.H.M. (1974), 'Taxation in antiquity', in P. Brunt (ed.), *The Roman Economy: Studies in Ancient Economic and Administrative History*, Oxford: 151-86.
- Jones, D. (2006), *The Bankers of Puteoli: Financing Trade and Industry in the Roman World*, Stroud.
- Jones, A.H.M. (1974), *The Roman Economy*, Oxford.
- Jones, G. (1980), 'The Roman mines at Riotinto', *Journal of Roman Studies* 70: 146-65.
- Jongman, W. (1988), *The Economy and Society of Pompeii*, Amsterdam.
- Jongman, W. (2006), 'The rise and fall of the Roman economy: population, rents and entitlements', in P.F. Bang, M. Ikeguchi and H.G. Ziche (eds.), *Ancient Economies, Modern Methodologies* Bari: 237-54.
- Jongman, W. (2007a), 'Gibbon was right: the decline and fall of the Roman economy', in O. Hekster, G. De Kleij and D. Slootjes (eds.), *Crisis and the Roman Empire*, Leiden: 183-99.
- Jongman, W. (2007b), 'The early Roman empire: Consumption', in W. Scheidel, I. Morris and R. Saller (eds.), *The Cambridge Economic History of the Greco-Roman World*, Cambridge: 592-618.
- Joshel, S. (1992), *Work, Identity, and Legal Status at Rome: A Study of the Occupational Inscriptions*, London.
- Kay, P. (2014), *Rome's Economic Revolution*, Oxford.
- Keay, J. (1993), *The Honourable Company: A History of the East India Company* 2 edn., London.
- Keay, S. (2009), *Scientific Approaches to the Study of Roman Ports* London.

- Kehoe, D. (1988a), 'Allocation of risk and investment on the estates of Pliny the Younger', *Chiron* 18: 15-42
- Kehoe, D. (1988b), *Economics of Agriculture on Roman Imperial Estates in North Africa*, Göttingen.
- Kehoe, D. (1997), *Investment, Profit and Tenancy. The Jurist and the Roman Economy*, Michigan.
- Kehoe, D. (2007), 'The early Roman empire: Production', in W. Scheidel, I. Morris and R. Saller (eds.), *The Cambridge Economic History of the Greco-Roman World*, Cambridge: 543-69.
- Keppie, L. (2005), 'The army and the navy', in A. Bowman, E. Champlin and A. Lintott (eds.), *The Cambridge Ancient History X: The Augustan Empire, 43 B.C.-A.D. 69* 2nd edn., Cambridge: 371-96.
- Kessler, D. and Temin, P. (2007), 'The organization of the grain trade in the early Roman empire', *The Economic History Review* 60 2: 313-32.
- Kron, G. (2008), 'The much maligned peasant. Comparative perspectives on the productivity of the small farmer in classical antiquity', in L. de Ligt and S. Northwood (eds.), *People, Land and Politics: Demographic Developments and the Transformation of Roman Italy, 300 BC- AD 1*, Leiden: 71-120.
- Kron, G. (2011), 'Food production', in W. Scheidel (ed.), *The Cambridge Companion to the Roman Economy*, Cambridge: 156-74.
- Krugman, P. and R. Wells (2013), *Economy*, 3rd edn., New York.
- Kucharik, C. and N. Ramankutty (2005), 'Trends and variability in U.S. corn yields over the twentieth century', *Earth Interactions* 9: 1-29.
- Laurence, R. (1999), *The Roads of Roman Italy*. London.
- Laurence, R., S. Cleary and G. Sears (2011), *The City in the Roman West c. 250 BC-c. AD 250*. Cambridge.
- Lawton, R. (1968), 'Population changes in England and Wales in the later nineteenth century: an analysis of trends by registration district', *Transactions of the Institute of British Geographers* 44: 55-74.
- London, J.E. (1990), 'The face on the coins and inflation in Roman Egypt', *Klio* 72: 106-34.
- Lo Cascio, E. (1981), 'State and coinage in the late Republic and early Empire', *Journal of Roman Studies* 71: 76-86.

- Lo Cascio, E. (2007), 'The early Roman Empire: the state and the economy', in W. Scheidel, I. Morris and R. Saller (eds.), *The Cambridge Economic History of the Greco-Roman World*, Cambridge: 619-47.
- Lo Cascio E. and P. Malanima (2005), 'Cycles and stability: Italian population before the demographic transition (225 B.C. – A.D. 1900)', *Rivista di Storia Economica* 21: 197-232.
- Lockyear, K. (1999), 'Hoard structure and coin production in antiquity – an empirical investigation', *The Numismatic Chronicle* 159: 215-43.
- MacKenzie, D. and G. Spinardi (1995), 'Tacit knowledge, weapons design, and the uninvention of nuclear weapons'. *The American Journal of Sociology* 101 1: 44-99.
- MacMullen, R. (1984), 'The Roman Emperor's army costs', *Latomus* 43: 571-80.
- Maddison, A. (2007), *Contours of the World Economy, 1-2030 A.D.: Essays in Macro-Economic History* Oxford.
- Mattingley, H. and Sydenham, E. (1923), *The Roman Imperial Coinage*. London.
- Mattingly, D. (1997), 'Beyond belief: drawing a line beneath the consumer city', in H. Parkins (ed.), *Roman Urbanisation: Beyond the Consumer City*. London: 210-8.
- Mayer, E. (2012), *The Ancient Middle Classes: Urban Life and Aesthetics in the Roman Empire, 100BCE-250 CE*. Cambridge.
- Mayerson, P. (1984), 'Wheat in the Roman world: an addendum', *The Classical Quarterly*: 243-5.
- Meiggs, R. (1973), *Roman Ostia*, 2 edn., Oxford.
- Mitchell, B. (2007), *International Historical Statistics 1750-2005: Europe*, 6th edn., New York.
- Mommsen, T., P. Krueger and A. Watson (1998) (eds.), *The Digest of Justinian*, Pennsylvania.
- Morley, N. (1996), *Metropolis and Hinterland*, Cambridge.
- Morley, N. (2000), 'Trajan's engines', *Greece and Rome* 47: 197-210.
- Morley, N. (2007), 'The early Roman empire: Distribution', in W. Scheidel, I. Morris and R. Saller (eds.), *The Cambridge Economic History of the Greco-Roman World*, Cambridge: 570-91.
- Morley, N. (2011), 'Cities and Economic Development in the Roman Empire' in A. Bowman and A. Wilson (eds.), *Settlement, Urbanisation and Population*, Oxford: 143-60.
- Neesen, L. (1980), *Untersuchungen zu den direkten Staatsabgaben der römischen Kaiserzeit (27 v. Chr.-284 n.Chr.)*, Bonn: 135-140.

- Nissen, J., P. Damerow and R. Englund (1993), *Archaic Bookkeeping: Early Writing and Techniques of Economic Administration in the Ancient Near East*, Chicago.
- North, D. (1990), *Institutions, Institutional Change and Economic Performance*, Cambridge.
- Parker, A. (1980), 'Ancient shipwrecks in the Mediterranean and the Roman provinces', *British Archaeological Reports, Supplementary Series*.
- Parker, A. (1990), 'Classical antiquity: the maritime dimension', *Antiquity* 64: 335-46.
- Parker, A. (2008), 'Artifact distribution and wreck locations: the archaeology of Roman commerce', in R. Hohlfelder (ed.), *The Maritime World of Ancient Rome*, Michigan: 177-98.
- Parkin, T. (1992), *Demography and Roman Society*, London.
- Parkins, H. (1997), 'The "consumer city" domesticated? The Roman city in elite economic strategies', in H. Parkins (ed.), *Roman Urbanisation: Beyond the Consumer City*, London: 83-7.
- Patten, J. (1978), *English Towns 1500-1700*, Kent.
- Pitts, M. and R. Griffin (2012), 'Exploring health and social well-being in late Roman Britain: An intercemetery approach', *American Journal of Archaeology* 116 2: 253-76.
- Polanyi, K. (1944), *The Great Transformation*, Boston.
- Pope, C. (1992), 'Adult Mortality in America before 1900: A View from Family Histories', in C. Goldin and H. Rockoff, (eds.), *Strategic Factors in Nineteenth Century American Economic History: A Volume to Honor Robert W. Fogel*, Chicago: 267-96.
- Population Investigation Committee (1964), 'The demography of the British peerage', *Population Studies Supplement*.
- Purcell, N. (1985), 'Wine and wealth in Roman Italy', *Journal of Roman Studies*:1-19.
- Rathbone, D. (1991), *Economic Rationalism and Rural Society in Third-Century A.D. Egypt: the Heroninos Archive and the Appianus Estate*, Cambridge.
- Rathbone, D. (1996a), 'Monetisation, not price-inflation, in third-century A.D. Egypt?', in C. King and D. Wigg (eds.), *Coin Finds and Coin Use in the Roman World*, Berlin: 321-39.
- Rathbone, D. (1981), 'The development of agriculture in the 'Ager Cosanus' during the Roamn Republic: Problems of evidence and interpretation', *Journal of Roman Studies*: 10-23.

- Rathbone, D. (1996b), 'The imperial finances', in A. Bowman, E. Champlin and A. Lintott (eds.), *The Cambridge Ancient History X: The Augustan Empire, 43 B.C.-A.D. 69*, 2nd edn., Cambridge: 309-23.
- Rathbone, D. (1997), 'Price and price formation in Roman Egypt', in A. Andreau, P. Briant and R. Descat (eds.), *Economie Antique: Prix et Formations des Prix dans les Economies Antiques*, St-Bertrand-de-Comminges: 183-244.
- Rathbone, D. (2000), 'The "Muziris" papyrus (SB XVIII 13167): financing Roman trade with India', *Bulletin de la Société d'Archéologie d'Alexandrie* 46: 39-50.
- Rathbone, D. (2003), 'The financing of maritime commerce in the Roman Empire, I-II AD', in E. Lo Cascio (ed.), *Credito e moneta nel mondo romano* Bari: 197-229.
- Rathbone, D. (2007a), 'Military finance and supply', in P. Sabin, H. Van Wees and M. Whitby (eds.), *The Cambridge History of Greek and Roman Warfare* Cambridge: 158-77.
- Rathbone, D. (2007b), 'Roman Egypt', in W. Scheidel, I. Morris and R. Saller (eds.), *The Cambridge Economic History of the Greco-Roman World*, Cambridge: 698-719.
- Rathbone, D. (2008), 'Poor peasants and silent sherds', in L. Ligt and S. Northwood (eds.), *People, Land and Politics: Demographic Developments and the Transformation of Roman Italy, 300 BC- A.D.*, Leiden: 305-32.
- Rathbone, D. (2009a), 'Earnings and costs: living standards and the Roman economy (first to third centuries A.D.)', in A. Bowman and A. Wilson (eds.), *Quantifying the Roman Economy: Methods and Problems*, Oxford: 299-326.
- Rathbone, D. (2009b), 'Nero's reforms of *vectigalia* and the inscription of the Lex Portorii Asiae', in M. Cottier, M. Crawford, C. Crowther, J. Ferrary, B. Levick, O. Salomies and M. Wörrle (eds.), *The Customs Law of Asia*, Oxford: 251-78.
- Rathbone, D. and P. Temin (2008), 'Financial intermediation in first-century AD Rome and eighteenth-century England', in K. Verboven, K. Vandorpe and V. Chankowski (eds.), *Pistoi dia tèn technèn. Bankers, Loans and Archives in the Ancient World. Studies in Honour of Raymond Bogaert*, Leuven: 371-419.
- Rathbone, D. and S. van Reden (2015), 'Mediterranean grain prices in classical antiquity', in R. van der Speck, B. van Leeuwen and J. van Zanden (eds.), *A History of Market Performance: From Ancient Babylonia to the Modern World*, London: 149-235.
- Rickman, G. (1971), *Roman Granaries and Store Buildings*, Cambridge.
- Rickman, G. (1980), *The Corn Supply of Ancient Rome*, Oxford.

- Rihill, T. (2013), *Technology and Society in the Ancient Greek and Roman Worlds*, American Historical Association.
- Robinson D. and A. Wilson (2011), 'Introduction: maritime archaeology and the ancient economy', in D. Robinson and A. Wilson (eds.), *Maritime Archaeology and Ancient Trade in the Mediterranean*, Oxford: 1-11.
- Roby, H. (1886), *De usufructu : Iustiniani Digestorum Lib. VII Tit. I*, Cambridge: 188-91.
- Romer, P. (1990), 'Endogenous technological change', *Journal of Political Economy* 98 5: 71-102.
- Rosenstein, N. (2004), *Rome at War: Farms, Families and Death in the Middle Republic*, North Carolina.
- Rosenstein, N. (2008), 'Aristocrats and agriculture in the Middle and Late Republic', *Journal of Roman Studies* 98: 1-26.
- Rostovtzeff, M. (1957), *The Social and Economic History of the Roman Empire*, 2nd edn., Oxford.
- Roth, U. (2007), *Thinking Tools, Food for Thought: Agricultural Slavery between Evidence and Models*, London.
- Ruffing, K. (2008), *Die Berufliche Spezialisierung in Handel und Handwerk. Untersuchungen zu ihrer Bedingungen in der Römischen Kaiserzeit im Östlichen Mittelmeerraum auf der Grundlage Griechischer Inschriften und Papyri*, Rahden.
- Russel, B. (2009), 'The dynamics of stone transport between the Roman Mediterranean and its hinterland', *Facta: A Journal of Roman Material Cultural Studies* 2: 107-26.
- Russell, C. (1948), *British Medieval Population*, Albuquerque.
- Sallares, R. (1991), *The Ecology of the Ancient Greek World*, Cornell.
- Saller, R. (1994), *Patriarchy, property and death in the Roman family*, Cambridge.
- Saller, R. (2012), 'Human capital and economic growth', in W. Scheidel (ed.), *The Cambridge Companion to the Roman Economy*, Cambridge: 71-88.
- Scheidel, W. (2001a), *Death on the Nile – Disease and the Demography of Roman Egypt*, Leiden.
- Scheidel, W. (2001b), *Debating Roman Demography*, Leiden.
- Scheidel, W. (2007), 'Disease and death', in P. Erdkamp (ed.), *The Cambridge Companion to Ancient Rome*, Cambridge: 45-59.
- Scheidel, W. (2011), 'A comparative perspective on the determinants of the scale and productivity of Roman maritime trade in the Mediterranean', in W. Harris and K. Iara

- (eds.), *Maritime Technology in the Ancient Economy: Ship-Design and Navigation*, Journal of Roman Archaeology Supplement 84: 21-37.
- Scheidel, W. and S. Friesen (2009), 'The size of the economy and the distribution of income in the Roman empire' *Journal of Roman Studies* 99: 66-91.
- Scott, S. (1932), *The Civil Law*, Cincinnati.
- Scott, W. (1910), *Money and Banking*, New York.
- Sealey, P. (2009), 'New light on the wine trade with Julio-Claudian Britain', *Britannia* 40: 1-40.
- Shaw, B. (2008), 'After Rome: transformation of the early Mediterranean world', *New Left Review* 51 89-114.
- Silver, M. (2008), 'The rise, demise, and (partial) rehabilitation of the peasant in Hopkins' model of Roman trade and taxes', *Classics Ireland* 15: 1-33.
- Slicher van Bath, B. (1963), *Yield ratios, 810-1820*, Wageningen.
- Speidel, M.A. (1992), 'Roman army pay scales', *Journal of Roman Studies* 82: 87-106.
- Speidel, M.P. (1970), 'The captor of Decebalus, a new inscription from Philippi', *Journal of Roman Studies* 60: 142-53.
- Speidel, M.P. (1973), 'The pay of the *auxilia*', *Journal of Roman Studies* 63: 141-7.
- Sperber, D. (1991), *Roman Palestine, 200-400: Money and Prices*, 2nd edn., Ramat-Gan.
- Spurr, M. (1986), *Arable Cultivation in Roman Italy*, London.
- Stein, P. (1962), 'Generations, life-spans and usufructs', *Revue Internationale des Droits de l'Antiquité* 9: 335-56.
- Stewart, W., D. Dibb, A. Johnston and T. Smyth (2005), 'The contribution of commercial fertilizer nutrients to food production', *Agronomy Journal* 97 1: 1-6.
- Stone, D. (2014), 'Africa in the Roman Empire: Connectivity, the economy and artificial port structures', *American Journal of Archaeology* 118 4: 565-600.
- Szaivert, W. and R. Wolters (2005), *Löhne, Preise, Werte. Quellen zur Römischen Geldwirtschaft*, Darmstadt.
- Szilágy, J. (1961), 'Beiträge zur Statistik der Sterblichkeit in den Westeuropäischen Provinzen des Römischen Imperiums', *Acta Archaeologica Academiae Scientiarum Hungaricae* 13: 125-55.
- Szilágy, J. (1962), 'Beiträge zur Statistik der Sterblichkeit in der Illyrischen Provinzgruppe und in Norditalien (Gallia Padana)', *Acta Archaeologica Academiae Scientiarum Hungaricae* 14: 297-396.

- Szilágy, J. (1963), 'Die Sterblichkeit in den Städten Mittel- und Süd-Italiens sowie in Hispanien (in den Römischen Kaiserzeit)', *Acta Archaeologica Academiae Scientiarum Hungaricae* 15: 129-224.
- Szilágy, J. (1965), 'Die Sterblichkeit in den Nord Afrikanischen Provinzen III', *Acta Archaeologica Academiae Scientiarum Hungaricae* 17: 309-34.
- Szilágy, J. (1966), 'Die Sterblichkeit in den Nord Afrikanischen Provinzen III', *Acta Archaeologica Academiae Scientiarum Hungaricae* 18: 235-77.
- Szilágy, J. (1967), 'Die Sterblichkeit in den Nord Afrikanischen Provinzen III', *Acta Archaeologica Academiae Scientiarum Hungaricae* 19: 25-59.
- Tchernia, A. (1983), 'Italian wines in Gaul at the end of the Republic', in P. Garnsey, K. Hopkins and C. Whittaker (eds.), *Trade in the Ancient Economy*, London: 87-105.
- Tchernia, A. (1988), *Le Vin de l'Italie Romaine*, Rome.
- Tchernia, A. (2011), *Les Romains et le Commerce*, Naples.
- Temin, P. (2004), 'Financial intermediation in the early Roman empire', *Journal of Economic History* 64 3: 705-33.
- Temin, P. (2006a), 'Estimating GDP in the early Roman empire', in E. Lo Cascio (ed.), *Innovazione tecnica e progresso economico nel mondo romano*, Bari: 31-54
- Temin, P. (2006b), 'The economy of the early Roman empire', *Journal of Economic Perspectives* 20 1:133-51.
- Temin, P. (2012), *Roman Market Economy*, Princeton.
- Thornton, M. And R. Thornton (1989), *Julio-Claudian Building Programs Wauconda*.
- Tillott, P. (ed.) (1961), *A History of the County of York: City of York (The Victoria History of the Counties of England)*, Oxford.
- Trenerry, C. (1926), *The Origin and Early History of Insurance*, London.
- Turner, P. (1989), *Roman Coins from India*, London.
- van Nijf, O. (1997), *The Civic World of Professional Associations in the Roman East*, Amsterdam.
- Walker, D. (1976), *The Metrology of the Roman Silver Coinage Part I From Augustus to Domitian*, Oxford.
- Walker, D. (1977), *The Metrology of the Roman Silver Coinage Part II From Nero to Commodus*, Oxford.
- Walker, D. (1978), *The Metrology of the Roman Silver Coinage Part III From Pertinax to Uranius Antoninus*, Oxford.
- Walsh, P. (1997), *Petronius. The Satyricon*, Oxford.

- Wassink, A. (1991), 'Inflation and financial policy under the Roman empire to the Price Edict of 301 A.D.', *Historia* 40: 465-92.
- Watson, G. (1956), 'The pay of the Roman army: Suetonius, Dio and the *quantum stipendium*', *Historia* 5: 332-40.
- Weber, M. (1930), *The Protestant Ethic and the Spirit of Capitalism*, London.
- Weber, M. (1978), *Economy and Society*, California.
- West L. (1929), *Imperial Roman Spain: The Objects of Trade*, Oxford.
- Whelan, K. (2012), 'ELA, promissory notes and all that: The fiscal cost of the Anglo-Irish Bank', *University College Dublin Centre for Economic Research Working Papers Series*: 12/06.
- White, K. (1963), 'Wheat-farming in Roman times', *Antiquity* 37: 207-12.
- White, K. (1965), 'The productivity of labour in Roman agriculture', *Antiquity* 39: 102-7.
- Whittaker, C. (1990), 'The consumer city revisited – the vicus and the city', *Journal of Roman Archaeology* 3: 110-8.
- Wickham, C. (2004), *Framing the Early Middle Ages: Europe and the Mediterranean, 400-800*, Oxford.
- Wilson, A. (2002), 'Machines, power and the ancient economy', *Journal of Roman Studies* 92: 1-32.
- Wilson, A. (2009), 'Approaches to quantifying Roman trade', in A. Bowman and A. Wilson (eds.), *Quantifying the Roman Economy: Methods and Problems*, Oxford: 213-49.
- Wilson, A. (2011), 'City sizes and urbanisation in the Roman Empire', in A. Bowman and A. Wilson (eds.), *Settlement, Urbanisation and Population*, Oxford: 161-95.
- Wilson, S. (1996), 'Voluntary associations: an overview', in J. Kloppenborg and S. Wilson (eds.), *Voluntary Associations in the Graeco-Roman World* New York: 1-15.
- Woods, R. (1993), 'On the historical relationship between infant and adult mortality', *Population Studies* 47 2: 195-219.
- Woolf, G. (2009), 'Literacy or literacies in Rome?', in W. Johnson and H. Parker (eds.), *Ancient Literacies: The Culture of Reading in Greece and Rome*, Oxford: 46-68.
- Zhao, Z. (1997), 'Long-term morality patterns in Chinese history: evidence from a recorded clan population', *Population Studies: A Journal of Demography* 51 2: 117-27.
- Zimmer, G. (1982), *Römische Berufsdarstellungen*, Berlin.